```
;       CITY:  King of Prussia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19406-2799
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/767,096
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Schreck, Patricia A.
;       REGISTRATION NUMBER:  33,777
;       REFERENCE/DOCKET NUMBER:  ATG50033
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5031
;       TELEFAX:  610-270-5090
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  284 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-767-096-3

  Query Match             7.2%;  Score 72.5;  DB 2;  Length 284;
  Best Local Similarity   23.8%;  Pred. No. 1.3;
  Matches   35;  Conservative   12;  Mismatches   35;  Indels   65;  Gaps    6;

Qy      12 PASISELLD-----CGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS 66
           | :|:|||||          || : ||       |:|: : : |       | |  ||||:
Db      88 PFNITELLDNIVSLTTAESESFVLDF-------PQPSADYLDFRNRLQTNHV-CLENCVL 139

Qy      67 KSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVC 126
           | |                                 : | |   ||  | | |
Db     140 KEK-------------------------------AIAGTVKVQNLAFEKVVK------- 160

Qy     127 DSSVVPTFELTLVFKQENCSWTSFRDF 153
             :  ||:          :| || ||
Db     161 ---IRMTFD----------TWKSFTDF 174


RESULT 9
US-09-480-203-3
; Sequence 3, Application US/09480203
; Patent No. 6297359
; GENERAL INFORMATION:
;  APPLICANT: Patricia T.W. Cohen
;  APPLICANT: Phillip Cohen
;  APPLICANT: Peter R. Young
;  TITLE OF INVENTION: A PROTEIN PHOSPHATASE 1 BINDING PROTEIN
;  TITLE OF INVENTION: R5
;  FILE REFERENCE: ATG-50033-2
;  CURRENT APPLICATION NUMBER: US/09/480,203
;  CURRENT FILING DATE: 2000-01-10
;  EARLIER APPLICATION NUMBER: 09/300,327
;  EARLIER FILING DATE: 1999-04-27
;  EARLIER APPLICATION NUMBER: 08/767,096
;  EARLIER FILING DATE: 1996-12-05
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-480-203-3

  Query Match             7.2%;  Score 72.5;  DB 4;  Length 284;
  Best Local Similarity   23.8%;  Pred. No. 1.3;
  Matches   35;  Conservative   12;  Mismatches   35;  Indels   65;  Gaps    6;

Qy      12 PASISELLD-----CGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS 66
           | :|:|||||          || : ||       |:|: : : |       | |  ||||:
Db      88 PFNITELLDNIVSLTTAESESFVLDF-------PQPSADYLDFRNRLQTNHV-CLENCVL 139

Qy      67 KSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVC 126
           | |                                 : | |   ||  | | |
Db     140 KEK-------------------------------AIAGTVKVQNLAFEKVVK------- 160

Qy     127 DSSVVPTFELTLVFKQENCSWTSFRDF 153
             :  ||:          :| || ||
Db     161 ---IRMTFD----------TWKSFTDF 174


RESULT 10
US-09-085-199B-9
; Sequence 9, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larson, Marina T.
;      REGISTRATION NUMBER:  32038
;      REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 668-2050
;      TELEFAX:  (970) 668-2052
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  756
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;    FEATURE:
;      OTHER INFORMATION:  Huntington-interacting protein
US-09-085-199B-9

  Query Match             7.2%;  Score 72;  DB 4;  Length 756;
  Best Local Similarity   24.5%;  Pred. No. 6.4;
  Matches   49;  Conservative   22;  Mismatches   71;  Indels   58;  Gaps   10;

Qy      12 PASISELLDCGYHPESLLSDFDYWDYVV----------PEP--NLNEV----------- 47
           |  |||||    |  :||:     | |:          |||  :| |
Db     401 PEDISELL----HSITLLAHLT-GDTVIQGSATSLRAPPEPADSLTEACRQYGRETLAYL 455

Qy      48 --IFEESTCQNL-VKMLENCLSK-------------SKQTKLG-------CSKVLVPEK 83
```

Filed 05-27-1998

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 13:54:32 ; Search time 88 Seconds
(without alignments)
379.409 Million cell updates/sec

Title: US-09-881-636-2
Perfect score: 1002
Sequence: 1 MVATGSLSSKNPASISELLD.....FRLVKKKLYSLIGTTVIEGS 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1002	100.0	193	4	Q96D03	Q96d03 homo sapien
2	932	93.0	193	11	Q9CQV2	Q9cqv2 mus musculu
3	271	27.0	232	4	Q9H0S3	Q9h0s3 homo sapien
4	271	27.0	232	4	Q9NX09	Q9nx09 homo sapien
5	262	26.1	229	11	Q9D3F7	Q9d3f7 mus musculu
6	211.5	21.1	299	5	Q9NHN5	Q9nhn5 drosophila
7	211.5	21.1	299	5	Q9VTI8	Q9vti8 drosophila
8	193	19.3	280	5	Q9VTH4	Q9vth4 drosophila
9	188	18.8	280	5	Q9NHN4	Q9nhn4 drosophila
10	91	9.1	902	12	Q9QB81	Q9qb81 yaba monkey
11	86.5	8.6	1089	10	Q9FN97	Q9fn97 arabidopsis
12	85.5	8.5	1414	5	O76411	O76411 caenorhabdi
13	84	8.4	902	12	Q9DHL2	Q9dhl2 yaba-like d
14	82.5	8.2	2081	10	Q9LH98	Q9lh98 arabidopsis
15	82	8.2	673	5	O76649	O76649 caenorhabdi
16	82	8.2	1002	16	Q97FZ3	Q97fz3 clostridium
17	81.5	8.1	228	17	Q97YG5	Q97yg5 sulfolobus
18	80.5	8.0	498	4	Q9NPE8	Q9npe8 homo sapien
19	80.5	8.0	582	4	Q9NSP8	Q9nsp8 homo sapien
20	80	8.0	443	5	Q9U0X3	Q9u0x3 leishmania
21	80	8.0	927	4	O75170	O75170 homo sapien
22	79.5	7.9	606	5	O44940	O44940 drosophila
23	79.5	7.9	1704	5	Q95Z06	Q95z06 trypanosoma
24	79	7.9	787	10	Q9C507	Q9c507 arabidopsis
25	78.5	7.8	680	5	O15740	O15740 dictyosteli
26	78	7.8	249	2	Q93UD4	Q93ud4 candidatus
27	78	7.8	315	11	Q9JJV5	Q9jjv5 mus musculu
28	78	7.8	1134	12	Q99D22	Q99d22 bovine herp
29	77.5	7.7	817	10	Q9ZPG9	Q9zpg9 arabidopsis
30	77.5	7.7	1153	10	Q9SL22	Q9sl22 arabidopsis
31	77.5	7.7	1584	10	Q9MAG6	Q9mag6 arabidopsis
32	77	7.7	600	4	Q96MN7	Q96mn7 homo sapien
33	77	7.7	674	11	Q9DCX9	Q9dcx9 mus musculu
34	77	7.7	1095	10	Q9LUZ5	Q9luz5 arabidopsis
35	77	7.7	1837	10	Q9LU53	Q9lu53 arabidopsis
36	76.5	7.6	574	5	Q960X1	Q960x1 drosophila
37	76.5	7.6	840	5	Q9N3Q5	Q9n3q5 caenorhabdi
38	76.5	7.6	1262	5	Q9W110	Q9w110 drosophila
39	76.5	7.6	1309	10	Q9SXF0	Q9sxf0 arabidopsis
40	76.5	7.6	1466	10	Q94IU9	Q94iu9 arabidopsis
41	76	7.6	332	10	Q94LH5	Q94lh5 oryza sativ
42	76	7.6	1083	3	Q06108	Q06108 saccharomyc
43	76	7.6	1871	10	Q9CAG6	Q9cag6 arabidopsis
44	76	7.6	1894	10	O64795	O64795 arabidopsis
45	75.5	7.5	423	16	Q9CNX7	Q9cnx7 pasteurella

ALIGNMENTS

RESULT 1
Q96D03
ID Q96D03 PRELIMINARY; PRT; 193 AA.
AC Q96D03;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO RIKEN CDNA 1700037B15 GENE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC013592; AAH13592.1; -.
SQ SEQUENCE 193 AA; 21740 MW; 3BEEB3D0530C5BFF CRC64;

Query Match 100.0%; Score 1002; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 3.8e-96;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MVATGSLSSKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKM 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVATGSLSSKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKM 60

Qy      61 LENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKK 120

Qy     121 LDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKK 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKK 180

Qy     181 LYSLIGTTVIEGS 193
           |||||||||||||
Db     181 LYSLIGTTVIEGS 193
```

```
RESULT 2
Q9CQV2
ID   Q9CQV2      PRELIMINARY;      PRT;   193 AA.
AC   Q9CQV2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   1700108M02RIK PROTEIN (1700037B15RIK PROTEIN).
GN   1700108M02RIK OR 1700037B15RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK018944; BAB31490.1; -.
DR   EMBL; AK006617; BAB24676.1; -.
DR   MGD; MGI:1920534; 1700037B15Rik.
DR   MGD; MGI:1925887; 1700108M02Rik.
SQ   SEQUENCE   193 AA;  21597 MW;  53648D776C1E5B0D CRC64;

  Query Match             93.0%;  Score 932;  DB 11;  Length 193;
  Best Local Similarity   93.2%;  Pred. No. 7e-89;
  Matches  178;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MVATGSLSSKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKM 60
           |||||||||||||||||||| |||| |||||||||||||||||||:|||:||||||||
Db       1 MVATGSLSSKNPASISELLDGGYHPGSLLSDFDYWDYVVPEPNLNEVVFEETTCQNLVKM 60

Qy      61 LENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKK 120
           ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LENCLSRSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKK 120

Qy     121 LDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKK 180
           |||||||:|||||||||||||||:| ||| :||||||||||| :|||||||:||||||
Db     121 LDRIVCDASVVPTFELTLVFKQESCPWTSLKDFFFSRGRFSSGLKRTLILSSGYRLVKKK 180

Qy     181 LYSLIGTTVIE 191
           ||||||||||
Db     181 LYSLIGTTVIE 191


RESULT 3
Q9H0S3
ID   Q9H0S3      PRELIMINARY;      PRT;   232 AA.
AC   Q9H0S3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 25.4 KDA PROTEIN.
GN   DKFZP564O2071.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT   Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL   Genome Res. 11:422-435(2001).
DR   EMBL; AL136668; CAB66603.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   232 AA;  25354 MW;  774E941EAD908198 CRC64;

  Query Match             27.0%;  Score 271;  DB 4;  Length 232;
  Best Local Similarity   35.0%;  Pred. No. 3.9e-20;
  Matches   63;  Conservative   45;  Mismatches   66;  Indels    6;  Gaps    4;

Qy       6 SLSSKNPASISELLDCGYHPESLLSDFDYWDYV-VPEPNLNEVIFEESTCQNLVKMLENC 64
           || | :  |: :  : |: ||    |  | | | :|:  |     :|  | ||:::|:
Db      48 SLESSDCESL-DSSNSGFGPE---EDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQES 103

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           |::::      :::|:| :|  :: :::|||: :|||||| :: | :|    |  : ::
Db     104 LAQARLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQL 163

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             | |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::||||||
Db     164 ALDPSLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 223


RESULT 4
Q9NX09
ID   Q9NX09      PRELIMINARY;      PRT;   232 AA.
AC   Q9NX09;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ20500 FIS, CLONE KAT09159 (HYPOTHETICAL 25.4 KDA
DE   PROTEIN).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UTERUS, LEIOMYOSARCOMA;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UTERUS, LEIOMYOSARCOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000507; BAA91214.1; -.
DR   EMBL; BC015236; AAH15236.1; -.
DR   EMBL; BC007714; AAH07714.1; -.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   232 AA;  25370 MW;  774E941EBDD08198 CRC64;

  Query Match             27.0%;  Score 271;  DB 4;  Length 232;
  Best Local Similarity   35.0%;  Pred. No. 3.9e-20;
  Matches   63;  Conservative   45;  Mismatches   66;  Indels    6;  Gaps    4;

Qy       6 SLSSKNPASISELLDCGYHPESLLSDFDYWDYV-VPEPNLNEVIFEESTCQNLVKMLENC 64
           || | :  |: :  : |: ||    |  | | | :|:  |     :|  | ||:::|:
Db      48 SLESSDCESL-DSSNSGFGPE---EDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQES 103

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           |::::      :::|:| :|  :: :::|||: :|||||||: :: | :|    |  : ::
Db     104 LAQARLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQL 163

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             | |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::||||||
Db     164 ALDPSLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 223


RESULT 5
Q9D3F7
ID   Q9D3F7      PRELIMINARY;      PRT;   229 AA.
AC   Q9D3F7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   5830413E08RIK PROTEIN.
GN   5830413E08RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=THYMUS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK017926; BAB31006.1; -.
DR   MGD; MGI:1921997; 5830413E08Rik.
SQ   SEQUENCE   229 AA;  24871 MW;  FE7EA24D3FF63949 CRC64;

  Query Match             26.1%;  Score 262;  DB 11;  Length 229;
  Best Local Similarity   30.7%;  Pred. No. 3.3e-19;
  Matches   65;  Conservative   45;  Mismatches   68;  Indels   34;  Gaps    5;

Qy       3 ATGSLSSKNPA--------------------SISELLDC--------GYHPESLLSDFD 33
           :: | ||: ||                    :  |  ||        |: ||    |
Db      12 SSSSSSSRTPAADRPPRSAWGSAAREEGLDRCASLESSDCESLDSSNSGFGPE---EDSS 68

Qy      34 YWDYV-VPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDV 92
           | | | :|:  |     :|  | ||:::|:  ||:::      :::|:| :|  :: :::
Db      69 YLDGVSLPDFELLSDPEDEHLCANLMQLLQESLSQARLGSRRPARLLMPSQLVSQVGKEL 128

Qy      93 LRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRD 152
              |||: :||||||| :: | :|    |  : ::  | |:||||:|||| : ::  |   :
Db     129 LRLAYSEPCGLRGALLDVCVEQGKSCHSVAQLALDPSLVPTFQLTLVLRLDSRLWPKIQG 188

Qy     153 FFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             |       |: ::| ||:|||::||||||
Db     189 LLSSANSSLVPGYSQSLTLSTGFRVIKKKLYS 220


RESULT 6
Q9NHN5
ID   Q9NHN5      PRELIMINARY;      PRT;   299 AA.
AC   Q9NHN5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CHARYBDE.
GN   CHARYBDE OR CG7533.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chauvet S., Maurel-Zaffran C., Miassod R., Jullien N., Pradel J.,
RA   Aragnol D.;
RT   "Characterization of charybde and scylla, two paralogous target genes
RT   of Hox and cofactors proteins in Drosophila.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF221109; AAF59840.1; -.
DR   FlyBase; FBgn0036165; charybde.
SQ   SEQUENCE   299 AA;  32133 MW;  9D6560B8963315E0 CRC64;

  Query Match             21.1%;  Score 211.5;  DB 5;  Length 299;
  Best Local Similarity   35.9%;  Pred. No. 8e-14;
  Matches   51;  Conservative   24;  Mismatches   60;  Indels    7;  Gaps    3;

Qy      43 NLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCG 102
           ||:::|     |  : | : |:  |  :|:  | |::| :|  |||||| :::|:|  ||||
Db     162 NLDDV--SASAVRELSQQLQAQLRDAKRRHLACTEVTLPNDLTQRIAAEIIRMSEREPCG 219

Qy     103 LRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFS 161
           | | : :  | | |  |::     |   |  ||| |  :|:   |:|    |      |
Db     220 ERACTLFIEFESEPNKVKRIAYFKVDPDTVSIFELYLTLRQDKSGWSSLVPQFIKNLTRS 279

Qy     162 SGFRRTLILSSGFRLVKKKLYS 183
           :    |: :|  | | ||||||
Db     280 N----TINISPDFTLTKKKLYS 297


RESULT 7
Q9VTI8
ID   Q9VTI8      PRELIMINARY;      PRT;   299 AA.
AC   Q9VTI8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CHARYBDE PROTEIN (LD22381P).
GN   CHARYBDE OR CG7533.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```

```
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003544; AAF50060.2; -.
DR   EMBL; AY060367; AAL25406.1; -.
DR   FlyBase; FBgn0036165; charybde.
SQ   SEQUENCE   299 AA;  32179 MW;  CD224D3BFBF315E7 CRC64;

  Query Match            21.1%;  Score 211.5;  DB 5;  Length 299;
  Best Local Similarity  35.9%;  Pred. No. 8e-14;
  Matches   51;  Conservative   24;  Mismatches   60;  Indels    7;  Gaps    3;

Qy      43 NLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCG 102
           ||:::|    |  : | : |:  |  :|:  | |::| :|  |||||| :::|:|  ||||
Db     162 NLDDV--SASAVRELSQQLQAQLRDAKRRHLACTEVTLPNDLTQRIAAEIIRMSEREPCG 219

Qy     103 LRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFS 161
           | | : :  | | |  |::     |   |  ||| |  :|:   |:|    |       |
Db     220 ERACTLFIEFESEPNKVKRIAYFKVDPDTVSIFELYLTLRQDKSGWSSLVPQFIKNLTRS 279

Qy     162 SGFRRTLILSSGFRLVKKKLYS 183
           :    |: :|  | | ||||||
Db     280 N----TINISPDFTLTKKKLYS 297


RESULT 8
Q9VTH4
ID   Q9VTH4      PRELIMINARY;      PRT;   280 AA.
AC   Q9VTH4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CG7590 PROTEIN.
GN   SCYLLA OR CG7590.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003545; AAF50074.1; -.
DR   FlyBase; FBgn0041094; scylla.
SQ   SEQUENCE   280 AA;  30762 MW;  894841884CD28258 CRC64;

  Query Match            19.3%;  Score 193;  DB 5;  Length 280;
  Best Local Similarity  29.2%;  Pred. No. 6.1e-12;
  Matches   49;  Conservative   36;  Mismatches   61;  Indels   22;  Gaps    5;

Qy      30 SDFDYWDYVVPE----------PNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVL 79
           |: :|: |   |           | ::   || :    :::|:  |  :|   | |::|
Db      99 SNSNYYYYAADEEEGGSADYALSNYDKKAVEELS----LRLLDE-LRAAKSRHLTCTEVS 153

Qy      80 VPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTL 138
           :|  ||  :|::::|:|  || |:||| :::  | |    :::  |  ||  | |||: |
Db     154 LPCDLTPSVAREIIRVSEKEPRGIRGCTIYIEFEDEPKNSRRIASIKVDSDTVSTFEVYL 213

Qy     139 VFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIG 186
           :|::  |||         :|     ||: :|  : : | ||||  |
Db     214 TLRQDHRGWTSLLP------QFMKSLARTITISPEYTITKNKLYSADG 255


RESULT 9
Q9NHN4
ID   Q9NHN4      PRELIMINARY;      PRT;   280 AA.
AC   Q9NHN4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   SCYLLA.
GN   SCYLLA OR CG7590.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapcda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chauvet S., Maurel-Zaffran C., Miassod R., Jullien N., Pradel J.,
RA   Aragnol D.;
RT   "Characterization of charybde and scylla, two paralogous target genes
RT   of Hox and cofactors proteins in Drosophila.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF221110; AAF59841.1; -.
DR   FlyBase; FBgn0041094; scylla.
SQ   SEQUENCE   280 AA;  30772 MW;  8B27BC184CD14E97 CRC64;

  Query Match             18.8%;  Score 188;  DB 5;  Length 280;
  Best Local Similarity   28.6%;  Pred. No. 2e-11;
  Matches   48;  Conservative   36;  Mismatches   62;  Indels   22;  Gaps    5;

Qy      30 SDFDYWDYVVPE----------PNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVL 79
           |: :|: |   |           | ::   || :   :::|:  |  :|   | |::|
Db      99 SNSNYYYYAADEEEGGSADYALSNYDKKAVEELS----LRLLDE-LRAAKSRHLTCTEVS 153

Qy      80 VPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTL 138
           :|  ||  :|:::::|:|  || |:||| :::  | |    :::  |  |   | |||: |
Db     154 LPCDLTPSVAREIIRVSEKEPRGIRGCTIYIEFEDEPKNSRRIASIKVDPDTVSTFEVYL 213

Qy     139 VFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIG 186
             :|::  |||         :|     ||: :|  : : | ||||  |
Db     214 TLRQDHRGWTSLLP------QFMKSLARTITISPEYTITKNKLYSADG 255


RESULT 10
Q9QB81
ID   Q9QB81      PRELIMINARY;      PRT;   902 AA.
AC   Q9QB81;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   YB-B17L.
GN   YB-B17L.
OS   Yaba monkey tumor virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Yatapoxvirus.
OX   NCBI_TaxID=38804;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Amano H., Morikawa S., Ueda Y., Miyamura T.;
RT   "Nucleotide sequence of the central 50kbp region of Yaba virus DNA.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95248279; PubMed=7730796;
RA   Amano H., Ueda Y., Miyamura T.;
RT   "Identification and characterization of the thymidine kinase gene of
RT   Yaba virus.";
RL   J. Gen. Virol. 76:1109-1115(1995).
DR   EMBL; AB015885; BAA88814.1; -.
SQ   SEQUENCE   902 AA;  102862 MW;  ED4373030A2BCF31 CRC64;

  Query Match              9.1%;  Score 91;  DB 12;  Length 902;
  Best Local Similarity   21.6%;  Pred. No. 1;
  Matches   50;  Conservative   41;  Mismatches   73;  Indels   68;  Gaps   11;

Qy       1 MVATGSLSSKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTC------ 54
           |:|  | | :  : :|   |            || :|   :  |:::||  |||
Db     354 MLAKKSRSKELKSKLSLFWD----------GMDYQEY--RQKKLSDIIFYNSTCYVMGLY 401

Qy      55 --QNLV--KMLENCLSKSKQTKLGC------SKVLVPEKLTQRI-------AQDVLRLSS 97
             |:    || : :| ::      |       |   ||: : : :        :|  : ||
Db     402 NKNNITYCSMLSDIMSANEIPLRVCVLPRVVSGKTVPDIIAETLNSINSISRKDFPKKSS 461

Qy      98 TEPCGLRGCVMHVNLEIENVCK--KLDRIVCDSSVVPTFELTLVF-------------- 140
           : |       ||: |   |  |  :| |:| ::|     :  |:|
Db     462 SAP-------MHIGLSENNFMKFFQLLRLVTNTSPERAIKEVLMFYAGLKLNDEGSPHLI 514

Qy     141 KQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGF-------RLVKKKLYSLI 185
           |:|  |:  |    ||   |    ::::| |:         |: |: :|:::
Db     515 KKE--SYQDFSVLLFSSMGFKVSIKKSIIASNNHTFIIVRPRVTKQYIYNML 564


RESULT 11
Q9FN97
ID   Q9FN97      PRELIMINARY;      PRT;  1089 AA.
AC   Q9FN97;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TRANSPOSON PROTEIN-LIKE.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=98069011; PubMed=9405937;
RA   Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT   Sequence features of the regions of 1,044,062 bp covered by thirteen
RT   physically assigned P1 clones.";
RL   DNA Res. 4:291-300(1997).
DR   EMBL; AB006705; BAB09502.1; -.
DR   InterPro; IPR000150; Hypothet_cof.
SQ   SEQUENCE   1089 AA;  126629 MW;  17664F2489EEB14E CRC64;

  Query Match              8.6%;  Score 86.5;  DB 10;  Length 1089;
  Best Local Similarity   19.2%;  Pred. No. 3.7;
  Matches   41;  Conservative   33;  Mismatches   83;  Indels   57;  Gaps    6;

Qy       2 VATGSLSSKNPASISELLDCG-------------YHPESLLSDFDYWDYVVPEPNLNEV 47
           : ||          : : :|||               :| :|:| :  ||  :    ||: :
Db     506 ILTGEELYNEVCCLPKTVDCGGNHGRLEGYGKTHNWHKQSILWELSYWKDLKLRHNLDVM 565

Qy      48 IFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCV 107
             |::   | :| | |   |:|                        ||   | |  :
Db     566 HIEKNVLDNFIKTLLNVQGKTKDNIKS------------------RLDLQENCNRKD-- 604

Qy     108 MHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSS----- 162
           :|:  |              : :| | |   |:    |   :|  || |  ||
Db     605 LHLTPE------------GKAPIPKFRLKPDAKEIFLRWLE-KDVKFSDGYSSSLANCV 650

Qy     163 ---GFRRTLILSSGFRLVKKKLYSLIGTTVIEGS 193
             | : | : |    :: ::|: :    ::: |
Db     651 DLPGRKLTGMKSHDCHVLMQRLFPIAFAELMDKS 684


RESULT 12
O76411
ID   O76411      PRELIMINARY;      PRT;  1414 AA.
AC   O76411;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 160.1 KDA PROTEIN.
GN   T10H9.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
```

```
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Greco T., Bradshaw H., O'Brien D.;
RT   "The sequence of C. elegans cosmid T10H9.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF067949; AAC19236.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00192; LDLa; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
DR   PROSITE; PS50060; MAM_2; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Glycoprotein; Hypothetical protein; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE   1414 AA;  160101 MW;  43800DFE3D48ACD4 CRC64;

  Query Match             8.5%;  Score 85.5;  DB 5;  Length 1414;
  Best Local Similarity   23.6%;  Pred. No. 6.4;
  Matches   46;  Conservative   35;  Mismatches   73;  Indels   41;  Gaps   10;

Qy      16 SELLDCGYHPESLLSDF---DY---WDYVVPEPNLNEVIFEESTCQNLVKM--------- 60
           |: :||| :    : ||   ||   |  |    : :| : | :|   | |:
Db     358 SDEMDCG-NINGTMCDFNGQDYCNSWYQVTNVTDYHERLSEPTTVAPLNKLNEVPLHLFR 416

Qy      61 LENCLSKSKQTKLGCSKVLV----PEKLTQRIAQDVL-RLSSTEPCG-------LRGCVM 108
           |::  :| |:   |   :||    |  ||:| :  |   |  | |          : | :
Db     417 LQSPSAKIKEAMRGSGNMLVFDHKPNPLTRRTSALVSPELPRTNPEAYDEKSPLFKSCKL 476

Qy     109 HVNL---------EIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGR 159
             |         :|  : | :: :    :::     ||: | |||:|      | :  |
Db     477 RFYLCSRTYSKVWQISVISKGINPMESGRTIIYEAGYTLIPK-ENCTW---ERVFVNIPR 532

Qy     160 FSSGFRRTLILSSGF 174
           ::|||  : ::: |
Db     533 QNAGFRIGIFVTNYF 547


RESULT 13
Q9DHL2
ID   Q9DHL2      PRELIMINARY;      PRT;   902 AA.
AC   Q9DHL2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   101L PROTEIN.
GN   101L.
OS   Yaba-like disease virus (YLDV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Yatapoxvirus.
OX   NCBI_TaxID=132475;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lee H.J.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21176366; PubMed=11277691;
RA   Lee H.J., Essani K., Smith G.L.;
RT   "The genome sequence of Yaba-like disease virus, a yatapoxvirus.";
RL   Virology 281:170-192(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lee H.J.;
RL   Thesis (2000), Sir William Dunn School of Pathology, University of.
DR   EMBL; AJ293568; CAC21339.1; -.
SQ   SEQUENCE   902 AA;  103042 MW;  F4A106217CD0CE05 CRC64;

  Query Match             8.4%;  Score 84;  DB 12;  Length 902;
  Best Local Similarity   20.0%;  Pred. No. 5.3;
  Matches   47;  Conservative   44;  Mismatches   70;  Indels   74;  Gaps   11;

Qy       1 MVATGSLSSKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTC------ 54
           |:|  | | :  :::|   |           || :|   :  |:::||  |||
Db     354 MLAKKSRSKELKSNLSLFWD----------GIDYQEY--RQKKLSDIIFYNSTCYVMGLY 401

Qy      55 --QNLV--KMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVL----------------R 94
             |:    || : :| ::     |   |:|  :: :   |::                 :
Db     402 NKNNITYCSMLSDIISFNETPLRVC---LLPRVVSGKTVPDIIAETLNSINSISRKDFPK 458

Qy      95 LSSTEPCGLRGCVMHVNLEIENVCK--KLDRIVCDSSVVPTFELTLVF------------ 140
             |:: |       ||: |   |  |  :| |:| ::       :  |:|
Db     459 KSASVP-------MHIGLTENNFMKFFQLLRLVTNTPPERAIKEVLMFYAGLKLNDDGSP 511

Qy     141 ---KQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGF-------RLVKKKLYSLI 185
             |:|  |:  |    ||   |    ::::| |:          |: |: :|:::
Db     512 HLIKKE--SYQDFSVLLFSAMGFKVSIKKSIIASNNHTFIIVRPRVTKQYIYNML 564


RESULT 14
Q9LH98
ID   Q9LH98      PRELIMINARY;      PRT;  2081 AA.
AC   Q9LH98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T19N8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP002057; BAB03174.1; -.
SQ   SEQUENCE   2081 AA;  232851 MW;  D3603E1F85EFFF29 CRC64;

  Query Match             8.2%;  Score 82.5;  DB 10;  Length 2081;
  Best Local Similarity   29.7%;  Pred. No. 21;
  Matches   30;  Conservative   13;  Mismatches   29;  Indels   29;  Gaps    5;

Qy     112 LEIENVC----------KKLDRIVCDSSVVP--------TFELTLVFKQENCSWTSFRD 152
```

```
          |::|| |           |  | :  :||| |        ||:  |  | ||   |
Db     51 LKLENNCPIREQLRSFFEKLKDLLKLESSVTPLIEDNEPKTFKFDLKSKSENLLQTM--- 107

Qy    153 FFFSRGRFSSGFRRTLILSSGFRLVKK--KLYSLIGTTVIE 191
          |   ||  ||  |: :     | ::|   :|:: ||  :||
Db    108 FMLGRGLLSSSIRKEM-----FEVMKSLTELHAAIGRVIIE 143


RESULT 15
O76649
ID   O76649      PRELIMINARY;      PRT;   673 AA.
AC   O76649;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 77.1 KDA PROTEIN.
GN   F25E5.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wilson R., Bradshaw H.;
RT   "The sequence of C. elegans cosmid F25E5.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF078157; AAG24078.1; -.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR001254; Trypsin.
DR   SMART; SM00225; BTB; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50097; BTB; 1.
KW   Hydrolase; Hypothetical protein; Serine protease.
SQ   SEQUENCE   673 AA;  77122 MW;  8ED29BE157C9D8C7 CRC64;

  Query Match             8.2%;  Score 82;  DB 5;  Length 673;
  Best Local Similarity   26.9%;  Pred. No. 6;
  Matches   29;  Conservative   21;  Mismatches   36;  Indels   22;  Gaps    5;

Qy      2 VATGSLSSKNPASISEL--------LDCGYHPESLLSDFDYWDYVVPEPNLN--EVIFEE 51
          :| | |:: :|  : ::        :|  |||:   | :   | :      |  ||:
Db    185 LARGRLNTADPYILEDITPIDFKLYMDLTYHPKQYFSAYHAKDILQIAKRFNNLEVV--- 241

Qy     52 STCQNLV--KMLENCLSKSKQTKLGCS-------KVLVPEKLTCRIAQ 90
          :||||::   : || ||   :|||  |        ::  |  :||  |:
Db    242 TTCQNVILEDLNENKLSTKNKTKLAESYDLDTVRELFNPNTMTCPSAR 289


Search completed: October 11, 2002, 14:13:46
Job time : 92 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 11, 2002, 14:11:18 ; Search time 32 Seconds
(without alignments)
147.317 Million cell updates/sec

Title: US-09-881-636-2
Perfect score: 1002
Sequence: 1 MVATGSLSSKNPASISELLD.....FRLVKKKLYSLIGTTVIEGS 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	8.4	286	2	US-08-576-626A-42	Sequence 42, Appl
2	79	7.9	549	5	PCT-US93-05701-2	Sequence 2, Appli
3	77	7.7	1417	2	US-08-559-303B-78	Sequence 78, Appl
4	77	7.7	1417	3	US-08-781-891-78	Sequence 78, Appl
5	77	7.7	1417	4	US-09-175-828-78	Sequence 78, Appl
6	74	7.4	650	1	US-08-325-071-59	Sequence 59, Appl
7	74	7.4	650	4	US-08-461-004A-59	Sequence 59, Appl
8	72.5	7.2	284	2	US-08-767-096-3	Sequence 3, Appli
9	72.5	7.2	284	4	US-09-480-203-3	Sequence 3, Appli
10	72	7.2	756	4	US-09-085-199B-9	Sequence 9, Appli
11	72	7.2	871	2	US-08-775-009-34	Sequence 34, Appl
12	70.5	7.0	530	4	US-09-180-852-2	Sequence 2, Appli
13	70.5	7.0	855	4	US-08-890-865A-10	Sequence 10, Appl
14	70	7.0	650	1	US-08-325-071-56	Sequence 56, Appl
15	70	7.0	650	4	US-08-461-004A-56	Sequence 56, Appl
16	70	7.0	688	1	US-08-325-071-57	Sequence 57, Appl
17	70	7.0	688	4	US-08-461-004A-57	Sequence 57, Appl
18	70	7.0	858	3	US-08-946-026-3	Sequence 3, Appli
19	70	7.0	1098	1	US-07-777-715-7	Sequence 7, Appli
20	70	7.0	1098	1	US-08-170-126-2	Sequence 2, Appli
21	70	7.0	1098	3	US-08-954-418-2	Sequence 2, Appli
22	69.5	6.9	905	4	US-09-360-186-3	Sequence 3, Appli
23	69	6.9	450	4	US-09-592-891A-14	Sequence 14, Appl
24	69	6.9	549	1	US-08-325-071-61	Sequence 61, Appl
25	69	6.9	549	4	US-08-461-004A-61	Sequence 61, Appl
26	69	6.9	650	1	US-08-325-071-63	Sequence 63, Appl
27	69	6.9	650	4	US-08-461-004A-63	Sequence 63, Appl
28	69	6.9	739	2	US-08-836-943-2	Sequence 2, Appli
29	68	6.8	280	1	US-08-471-058-19	Sequence 19, Appl
30	68	6.8	280	3	US-08-471-057-19	Sequence 19, Appl
31	68	6.8	280	4	US-09-234-186-3	Sequence 3, Appli
32	68	6.8	280	5	PCT-US93-05651-3	Sequence 3, Appli
33	68	6.8	481	3	US-08-959-381A-1	Sequence 1, Appli
34	68	6.8	542	3	US-08-959-381A-2	Sequence 2, Appli
35	68	6.8	620	1	US-08-325-071-65	Sequence 65, Appl
36	68	6.8	620	4	US-08-461-004A-65	Sequence 65, Appl
37	68	6.8	650	1	US-08-325-071-67	Sequence 67, Appl
38	68	6.8	650	4	US-08-461-004A-67	Sequence 67, Appl
39	68	6.8	3079	5	PCT-US94-00198-4	Sequence 4, Appli
40	67.5	6.7	613	1	US-08-465-687A-2	Sequence 2, Appli
41	67.5	6.7	613	4	US-09-030-970-2	Sequence 2, Appli
42	67.5	6.7	613	5	PCT-US94-11843-2	Sequence 2, Appli
43	67	6.7	257	3	US-08-981-256A-12	Sequence 12, Appl
44	66.5	6.6	1023	1	US-08-198-446B-4	Sequence 4, Appli
45	66.5	6.6	1023	2	US-08-870-693-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-576-626A-42
; Sequence 42, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-42

Query Match 8.4%; Score 84; DB 2; Length 286;
Best Local Similarity 22.2%; Pred. No. 0.051;

```
  Matches   35;  Conservative   24;  Mismatches   57;  Indels   42;  Gaps    6;

Qy      75 CSKVLVPE----------------------KLTQRIAQDVLRLSSTEPCGLRGCVMHVN 111
           |:|: :||                      |  : | : |  : | |    :  :  ::
Db       3 CTKLSIPEVILFEPRIFEDDRGHFFESFNLAKFQESIGRQVTFVQSNESYSKQNVIRGLH 62

Qy     112 LEIENVCKKLDRI----VCDSSV---------VPTFELTLVFKQ---ENCSWTSFRDFFF 155
            ::     || |:    | | :|         || :   ::||:   | |    |  |
Db      63 YQVIRPQGKLVRVVEGEVFDIAVDLRKSLGVIVPHY--LMIFKKLDIEGCYLIEFNKFID 120

Qy     156 SRGRFSSGFRRTLILSSGFRL-VKKKLYSLIGTTVIEG 192
           ||| |   |       :|  | :::: ||:    || |
Db     121 SRGTFVKTFHSDFFSENGIVLDMREEFYSISAKNVIRG 158


RESULT 2
PCT-US93-05701-2
; Sequence 2, Application PC/TUS9305701
;  GENERAL INFORMATION:
;    APPLICANT:  Massachusetts Institute of Technology
;    TITLE OF INVENTION:  Cloning and Characterization of Cell Death Genes
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Massachusetts Institute of Technology
;      STREET:  77 Massachusetts Avenue
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02139
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  diskette
;      COMPUTER:
;      OPERATING SYSTEM:
;      SOFTWARE:
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05701
;      FILING DATE:  19930614
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  549 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-05701-2

  Query Match             7.9%;  Score 79;  DB 5;  Length 549;
  Best Local Similarity   24.7%;  Pred. No. 0.55;
  Matches   45;  Conservative   23;  Mismatches   46;  Indels   68;  Gaps   11;

Qy      10 KNPASISELLD-CGYHPESLLSDF--DYWDYVVPEPNLNE--VIFEESTCQNLVKMLENC 64
           :  : :  |:|   |: :| |:||  || |: : ||:|    ||  : : | |
Db      63 RQASELGPLIDFFNYNNQSHLADFLEDYIDFAINEPDLLRPVVIAPQFSRQML------- 115

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
              ::  ||     ||:::|              | :|          :: | ||||
Db     116 ---DRKLLLG----NVPKQMT--------------CYIR------EYHVDRVIKKLD-- 145

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSL 184
                        | |   ||  | |  ||  ||  :::| |        |   |
Db     146 ------------------EMCDLDSF--FLFLHGRAGSG--KSVIASQAL----SKSDQL 179

Qy     185 IG 186
           ||
Db     180 IG 181


RESULT 3
US-08-559-303B-78
; Sequence 78, Application US/08559303B
; Patent No. 5824501
;  GENERAL INFORMATION:
;    APPLICANT:  NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
;    APPLICANT:  GRODEN
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND TREATMENT
;    TITLE OF INVENTION:  OF BLOOM'S SYNDROME
;    NUMBER OF SEQUENCES:  78
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,303B
;      FILING DATE:  NOVEMBER 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  63475/65
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1417
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OTHER NUCLEIC ACID
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-559-303B-78

  Query Match             7.7%;  Score 77;  DB 2;  Length 1417;
  Best Local Similarity   31.8%;  Pred. No. 4;
  Matches   34;  Conservative   21;  Mismatches   32;  Indels   20;  Gaps    8;

Qy      31 DFDY-WDYVVPEPNLNEVIFEEST---CQNLVKMLENCLSKSKQTKLGCSKVLV--PEKL 84
           | ||  |:| | |   |:|   |:   | : :| |:   |  |:  |  || |:  |||:
Db     293 DDDYDTDFVPPSP--EEIISASSSSSKCLSTLKDLDT--SDRKEDVLSTSKDLLSKPEKM 348

Qy      85 TQRIAQDVLRLSSTE----PCGLRGCVMHVNLEIENVCKKLDRIVCD 127
              :    |::   :|:       |:  ::||   :|::|| :| |  |
Db     349 SM---QELNPETSTDCDARQISLQQQLIHV---MEHICKLIDTIPDD 389


RESULT 4
US-08-781-891-78
; Sequence 78, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
```

```
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104-7092
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/781,891
;       FILING DATE:  27-DEC-1996
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 6090620tenburg Ph.D., Carol
;       REGISTRATION NUMBER:  39,317
;       REFERENCE/DOCKET NUMBER:  240052.419
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  78:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1417 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-781-891-78

  Query Match             7.7%;  Score 77;  DB 3;  Length 1417;
  Best Local Similarity   31.8%;  Pred. No. 4;
  Matches   34;  Conservative   21;  Mismatches   32;  Indels   20;  Gaps    8;

Qy      31 DFDY-WDYVVPEPNLNEVIFEEST---CQNLVKMLENCLSKSKQTKLGCSKVLV--PEKL 84
           | ||  |:| | |   |:|   |:   | : :| |:   |  |:  |  || |:  |||:
Db     293 DDDYDTDFVPPSP--EEIISASSSSSKCLSTLKDLDT--SDRKEDVLSTSKDLLSKPEKM 348

Qy      85 TQRIAQDVLRLSSTE----PCGLRGCVMHVNLEIENVCKKLDRIVCD 127
           :   |::   :||:       |:  ::||   :|::|| :| |  |
Db     349 SM---QELNPETSTDCDARQISLQQQLIHV---MEHICKLIDTIPDD 389


RESULT 5
US-09-175-828-78
; Sequence 78, Application US/09175828
; Patent No. 6221643
;  GENERAL INFORMATION:
;    APPLICANT:  NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
;    APPLICANT:  GRODEN
;    TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND TREATMENT
;    TITLE OF INVENTION:  OF BLOOM'S SYNDROME
;    NUMBER OF SEQUENCES:  78
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/175,828
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,303
;      FILING DATE:  NOVEMBER 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  63475/65
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1417
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OTHER NUCLEIC ACID
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-09-175-828-78

  Query Match             7.7%;  Score 77;  DB 4;  Length 1417;
  Best Local Similarity   31.8%;  Pred. No. 4;
  Matches   34;  Conservative   21;  Mismatches   32;  Indels   20;  Gaps    8;

Qy      31 DFDY-WDYVVPEPNLNEVIFEEST---CQNLVKMLENCLSKSKQTKLGCSKVLV--PEKL 84
           | ||  |:| | |   |:|   |:   | : :| |:   |  |:  |  || |:  |||:
Db     293 DDDYDTDFVPPSP--EEIISASSSSSKCLSTLKDLDT--SDRKEDVLSTSKDLLSKPEKM 348

Qy      85 TQRIAQDVLRLSSTE----PCGLRGCVMHVNLEIENVCKKLDRIVCD 127
           :   |::   :||:       |:  ::||   :|::|| :| |  |
Db     349 SM---QELNPETSTDCDARQISLQQQLIHV---MEHICKLIDTIPDD 389


RESULT 6
US-08-325-071-59
; Sequence 59, Application US/08325071
; Patent No. 5587311
;  GENERAL INFORMATION:
;    APPLICANT:  COBON, Stewart Gary
;    APPLICANT:  MOORE, Joanna Terry
;    APPLICANT:  JOHNSON, Law Anthony York
;    APPLICANT:  WILLADSEN, Peter
;    APPLICANT:  KEMP, David Harold
;    APPLICANT:  SRISKANTHA, Alagacone
;    APPLICANT:  RIDING, George Alfred
;    APPLICANT:  RAND, Keith No. 5587311man
;    TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;    TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,071
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,109
;      FILING DATE:  17-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/926,368
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/242,196
;      FILING DATE:  06-JUL-1988
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/AU87/00401
;       FILING DATE:  27-NOV-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI4912
;       FILING DATE:  16-OCT-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI2570
;       FILING DATE:  19-JUN-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PH9196
;       FILING DATE:  27-NOV-1986
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  60042/111 BIAU
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 672 5300
;       TELEFAX:  202 672 5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  650 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-325-071-59

  Query Match             7.4%;  Score 74;  DB 1;  Length 650;
  Best Local Similarity   24.4%;  Pred. No. 2.9;
  Matches   44;  Conservative   25;  Mismatches   47;  Indels   64;  Gaps   12;

Qy      17 ELLDCGYHPESLLSDFDYWDYVVPEPN-------LNEVIFEESTCQNL------VKMLEN 63
           | :||| :  :  | :  |    | ||       |||  :  |   |:       |  |:
Db     303 EFMDCGVY-MNRQSCYCPWKSRKPGPNVNINGCLLNEYYYTVSFTPNISFDSDHCKWYED 361

Qy      64 CLSKSKQTKLG----------CSK----VLVPEKLTQRIAQDVLR-LSSTE-PCGLRGCV 107
             : :: :| :|          |::     |: |||   ::: ||| | : | | |   |:
Db     362 RVLEAIRTSIGKEVFKVEILNCTQDIKARLIAEKL---LSKHVLRKLQACEHPIG-EWCM 417

Qy     108 MHVNL--------EIE------------------NVCKKLDRIV---CDSSVVPTFELT 137
           |:  |        |||                  | | |:| :    |      |:|:|
Db     418 MYPKLLIKKNSATEIEEENLCDSLLKDQEAAYKGQNKCVKVDNLFWFQCADGYTTTYEMT 477


RESULT 7
US-08-461-004A-59
; Sequence 59, Application US/08461004A
; Patent No. 6235283
;  GENERAL INFORMATION:
;     APPLICANT:  COBON, Stewart Gary
;     APPLICANT:  MOORE, Joanna Terry
;     APPLICANT:  JOHNSON, Law Anthony York
;     APPLICANT:  WILLADSEN, Peter
;     APPLICANT:  KEMP, David Harold
;     APPLICANT:  SRISKANTHA, Alagacone
;     APPLICANT:  RIDING, George Alfred
;     APPLICANT:  RAND, Keith No. 6235283man
;     TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;     TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;     NUMBER OF SEQUENCES:  71
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W.
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/461,004A
;       FILING DATE:  04-JUN-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/325,071
;       FILING DATE:  19-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/062,109
;       FILING DATE:  17-MAY-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/926,368
;       FILING DATE:  07-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/242,196
;       FILING DATE:  06-JUL-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/AU87/00401
;       FILING DATE:  27-NOV-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI4912
;       FILING DATE:  16-OCT-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI2570
;       FILING DATE:  19-JUN-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PH9196
;       FILING DATE:  27-NOV-1986
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  60042/152
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 672 5300
;       TELEFAX:  202 672 5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  650 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-461-004A-59

  Query Match             7.4%;  Score 74;  DB 4;  Length 650;
  Best Local Similarity   24.4%;  Pred. No. 2.9;
  Matches   44;  Conservative   25;  Mismatches   47;  Indels   64;  Gaps   12;

Qy      17 ELLDCGYHPESLLSDFDYWDYVVPEPN-------LNEVIFEESTCQNL------VKMLEN 63
           | :||| :  :  | :  |    | ||       |||  :  |   |:       |  |:
Db     303 EFMDCGVY-MNRQSCYCPWKSRKPGPNVNINGCLLNEYYYTVSFTPNISFDSDHCKWYED 361

Qy      64 CLSKSKQTKLG----------CSK----VLVPEKLTQRIAQDVLR-LSSTE-PCGLRGCV 107
             : :: :| :|          |::     |: |||   ::: ||| | : | | |   |:
Db     362 RVLEAIRTSIGKEVFKVEILNCTQDIKARLIAEKL---LSKHVLRKLQACEHPIG-EWCM 417

Qy     108 MHVNL--------EIE------------------NVCKKLDRIV---CDSSVVPTFELT 137
           |:  |        |||                  | | |:| :    |      |:|:|
Db     418 MYPKLLIKKNSATEIEEENLCDSLLKDQEAAYKGQNKCVKVDNLFWFQCADGYTTTYEMT 477


RESULT 8
US-08-767-096-3
; Sequence 3, Application US/08767096
; Patent No. 5939284
;  GENERAL INFORMATION:
;     APPLICANT:  Cohen, Patricia T.W.
;     APPLICANT:  Cohen, Philip
;     APPLICANT:  Young, Peter R.
;     TITLE OF INVENTION:  A Protein Phosphatase 1 Binding Protein,
;     TITLE OF INVENTION:  R5
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road
```

```
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-2799
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/767,096
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schreck, Patricia A.
;      REGISTRATION NUMBER:  33,777
;      REFERENCE/DOCKET NUMBER:  ATG50033
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5031
;      TELEFAX:  610-270-5090
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  284 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-767-096-3

  Query Match             7.2%;  Score 72.5;  DB 2;  Length 284;
  Best Local Similarity   23.8%;  Pred. No. 1.3;
  Matches   35;  Conservative   12;  Mismatches   35;  Indels   65;  Gaps    6;

Qy      12 PASISELLD-----CGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS 66
           | :|:|||          || : ||       |:|: : : |      | |  ||||:
Db      88 PFNITELLDNIVSLTTAESESFVLDF-------PQPSADYLDFRNRLQTNHV-CLENCVL 139

Qy      67 KSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVC 126
           | |                              : | |   ||  | | |
Db     140 KEK-------------------------------AIAGTVKVQNLAFEKVVK------- 160

Qy     127 DSSVVPTFELTLVFKQENCSWTSFRDF 153
             :  ||:          :| || ||
Db     161 ---IRMTFD----------TWKSFTDF 174


RESULT 9
US-09-480-203-3
; Sequence 3, Application US/09480203
; Patent No. 6297359
; GENERAL INFORMATION:
;  APPLICANT: Patricia T.W. Cohen
;  APPLICANT: Phillip Cohen
;  APPLICANT: Peter R. Young
;  TITLE OF INVENTION: A PROTEIN PHOSPHATASE 1 BINDING PROTEIN
;  TITLE OF INVENTION: R5
;  FILE REFERENCE: ATG-50033-2
;  CURRENT APPLICATION NUMBER: US/09/480,203
;  CURRENT FILING DATE: 2000-01-10
;  EARLIER APPLICATION NUMBER: 09/300,327
;  EARLIER FILING DATE: 1999-04-27
;  EARLIER APPLICATION NUMBER: 08/767,096
;  EARLIER FILING DATE: 1996-12-05
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-480-203-3

  Query Match             7.2%;  Score 72.5;  DB 4;  Length 284;
  Best Local Similarity   23.8%;  Pred. No. 1.3;
  Matches   35;  Conservative   12;  Mismatches   35;  Indels   65;  Gaps    6;

Qy      12 PASISELLD-----CGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS 66
           | :|:|||          || : ||       |:|: : : |      | |  ||||:
Db      88 PFNITELLDNIVSLTTAESESFVLDF-------PQPSADYLDFRNRLQTNHV-CLENCVL 139

Qy      67 KSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVC 126
           | |                              : | |   ||  | | |
Db     140 KEK-------------------------------AIAGTVKVQNLAFEKVVK------- 160

Qy     127 DSSVVPTFELTLVFKQENCSWTSFRDF 153
             :  ||:          :| || ||
Db     161 ---IRMTFD----------TWKSFTDF 174


RESULT 10
US-09-085-199B-9
; Sequence 9, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larson, Marina T.
;      REGISTRATION NUMBER:  32038
;      REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 668-2050
;      TELEFAX:  (970) 668-2052
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  756
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;    FEATURE:
;      OTHER INFORMATION:  Huntington-interacting protein
US-09-085-199B-9

  Query Match             7.2%;  Score 72;  DB 4;  Length 756;
  Best Local Similarity   24.5%;  Pred. No. 6.4;
  Matches   49;  Conservative   22;  Mismatches   71;  Indels   58;  Gaps   10;

Qy      12 PASISELLDCGYHPESLLSDFDYWDYVV----------PEP--NLNEV------------ 47
           |  |||||    |  :||:     | |:           |||  :| |
Db     401 PEDISELL----HSITLLAHLT-GDTVIQGSATSLRAPPEPADSLTEACRQYGRETLAYL 455

Qy      48 --IFEESTCQNL-VKMLENCLSK-------------SKQTKLG-------CSKVLVPEK 83
```

Filed 05-27-1998

```
           : || | :|  |  | ||||:               || :||        :     |
Db    456 SSLEEEGTVENADVTALRNCLSRVKTLGEELLPRGLDIKQEELGDLVDKEMAATSAAIEA 515

Qy     84 LTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQE 143
          | || :    : |    |  |  : ||  |   |  | : :    || : :|    :
Db    516 ATTRIEE----ILSKSRAGDTGVKLEVNERILGSCTSLMQAI-KVLVVASKDLQKEIVES 570

Qy    144 NCSWTSFRDFFFSRGRFSSG 163
               | ::|:    |:: |
Db    571 GRGSASPKEFYAKNSRWTEG 590


RESULT 11
US-08-775-009-34
; Sequence 34, Application US/08775009
; Patent No. 5935783
;  GENERAL INFORMATION:
;    APPLICANT:  Gong, Weilong
;    APPLICANT:  Emanuel, Beverly S.
;    APPLICANT:  Budarf, Marcia L.
;    APPLICANT:  Roe, Bruce
;    TITLE OF INVENTION:  No. 5935783el Genes Mapping in the Digeorge and
;    TITLE OF INVENTION:  Veliocardiofacial Syndrome Minimal Critical Region
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5935783ris, LLP
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/775,009
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trujillo, Doreen Yatko
;      REGISTRATION NUMBER:  35,719
;      REFERENCE/DOCKET NUMBER:  CH-0681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  871 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-775-009-34

  Query Match             7.2%;  Score 72;  DB 2;  Length 871;
  Best Local Similarity   27.2%;  Pred. No. 7.9;
  Matches   41;  Conservative   24;  Mismatches   52;  Indels   34;  Gaps    9;

Qy     10 KNPASISEL----LDCGYHPESLLSDFDYWDYVV-------PEPNLNEVIFEESTCQ-NL 57
          : ||: : |    :|    ||  | :  |:|  |         :|:|  | :|   |   |
Db    107 EEPATHNALAKIYIDSNNSPECFLRENAYYDSSVVGRYCEKRDPHLACVAYERGQCDLEL 166

Qy     58 VKML-ENCLSKSKQTKLGCSKVLVPE-------------KLTQRIAQDVLRLSSTEPCG 102
          :|:  || | ||:   | | |   ||              :|  :: | |   : :|
Db    167 IKVCNENSLFKSEARYLVCRK--DPELWAHVLEETNPSRRQLIDQVVQTALS-ETRDPEE 223

Qy    103 LRGCV---MHVNLEIENVCKKLDRIVCDSSV 130
          :   |   |  :|  | : : |::|| |:||
Db    224 ISVTVKAFMTADLPNE-LIELLEKIVLDNSV 253


RESULT 12
US-09-180-852-2
; Sequence 2, Application US/09180852
; Patent No. 6287834
; GENERAL INFORMATION:
;  APPLICANT: BELANGER, Alain
;  APPLICANT: HUM, Dean W.
;  APPLICANT: BEAULIEU, Martin
;  APPLICANT: LEVESQUE, Eric
;  TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
;  TITLE OF INVENTION: DIPHOSPHO-GLUCURONOSYLTRANSFERASE
;  FILE REFERENCE: 1259-449
;  CURRENT APPLICATION NUMBER: US/09/180,852
;  CURRENT FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: PCT/CA97/00328
;  EARLIER FILING DATE: 1997-05-16
;  EARLIER APPLICATION NUMBER: US 08/649,319
;  EARLIER FILING DATE: 1996-05-17
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 530
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-180-852-2

  Query Match             7.0%;  Score 70.5;  DB 4;  Length 530;
  Best Local Similarity   28.6%;  Pred. No. 5.7;
  Matches   34;  Conservative   12;  Mismatches   42;  Indels   31;  Gaps    6;

Qy      1 MVATGSLSSKNPASISELLDCGYHPESL----LSD-----FDYWDYVVPE-------PNL 44
            :| | | |    || |  :    :| ||    | |     || | | : :         |
Db     54 IVLTSSASILVNASKSSAIKLEVYPTSLTKNDLEDFFMKMFDRWTYSISKNTFWSYFSQL 113

Qy     45 NEVIFEESTCQNLVKMLENC-LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCG 102
           |: :| |     :|: |:  |:|    ||  ||             |||   :  |||
Db    114 QELCWEYSDYN--IKLCEDAVLNKKLMRKLQESKF-----------DVLLADAVNPCG 158


RESULT 13
US-08-890-865A-10
; Sequence 10, Application US/08890865A
; Patent No. 6307019
;  GENERAL INFORMATION:
;    APPLICANT:  Constantini, Franklin
;    APPLICANT:  Zeng, Li
;    TITLE OF INVENTION:  AXIN GENE AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,865A
;      FILING DATE:  10-JUL-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/54249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)278-0400
;      TELEFAX:  (212)391-0526
;  INFORMATION FOR SEQ ID NO:  10:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  855 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-890-865A-10

  Query Match             7.0%;  Score 70.5;  DB 4;  Length 855;
  Best Local Similarity   20.6%;  Pred. No. 12;
  Matches   35;  Conservative   28;  Mismatches   62;  Indels   45;  Gaps    6;

Qy       2 VATGSLSSKNPASISEL---------LDCGYHPE----------------SLLSDFDYW 35
           |: |  |||: |  :|          || || ||                ||| | |
Db      54 VSHGFYSSKSDAVRNETSTATPRRSDLDLGYEPEGSASPTPPYLKWAESLHSLLDDQD-- 111

Qy      36 DYVVPEPNLNEVIFEESTCQNLV---------KMLENCLSKSKQTKLGCSKVLVPEKLTQ 86
                ||     ::  | :|:          : || |:| ::: :|  :| :  : :
Db     112 -----GINLFRTFLKQEDCADLLDFWFACSGFRKLEPCVS-NEEKRLKLAKAIYKKYILD 165

Qy      87 RIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVC--DSSVVPTF 134
                ::       :: ||| : :: :   :    | |  : :  | |
Db     166 NNGIVSRQIKPATKSFIKDCVMKLQIDPDMFDQAQTEIQCMIEDNTYPLF 215


RESULT 14
US-08-325-071-56
; Sequence 56, Application US/08325071
; Patent No. 5587311
;  GENERAL INFORMATION:
;    APPLICANT:  COBON, Stewart Gary
;    APPLICANT:  MOORE, Joanna Terry
;    APPLICANT:  JOHNSON, Law Anthony York
;    APPLICANT:  WILLADSEN, Peter
;    APPLICANT:  KEMP, David Harold
;    APPLICANT:  SRISKANTHA, Alagacone
;    APPLICANT:  RIDING, George Alfred
;    APPLICANT:  RAND, Keith No. 5587311man
;    TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;    TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,071
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,109
;      FILING DATE:  17-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/926,368
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/242,196
;      FILING DATE:  06-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU87/00401
;      FILING DATE:  27-NOV-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PI4912
;      FILING DATE:  16-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PI2570
;      FILING DATE:  19-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PH9196
;      FILING DATE:  27-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  60042/111 BIAU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 672 5300
;      TELEFAX:  202 672 5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  650 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-071-56

  Query Match             7.0%;  Score 70;  DB 1;  Length 650;
  Best Local Similarity   22.6%;  Pred. No. 8.9;
  Matches   40;  Conservative   28;  Mismatches   51;  Indels   58;  Gaps   11;

Qy      17 ELLDCGYHPESLLSDFDYWDYVVPEPN-------LNEVIFEESTCQNL------VKMLEN 63
           | :||| :  :  | :  |    | ||       |||  :  |   |:       |  |:
Db     303 EFMDCGVY-MNRQSCYCPWKSRKPGPNVNINECLLNEYYYTVSFTPNISFDSDHCKWYED 361

Qy      64 CLSKSKQTKLG----------CSKVLVPEKLTQR-IAQDVLR-LSSTE-PCGLRGCVMHV 110
             : :: :| :|          |:: :    : :: ::: ||| | : | | |   |:|:
Db     362 RVLEAIRTSIGKEVFKVEILNCTQDIKARLIAEKPLSKHVLRKLQACEHPIG-EWCMMYP 420

Qy     111 NL--------EIE-----------------NVCKKLDRIV---CDSSVVPTFELT 137
           |         |||                 | | |:| :    |      |:|:|
Db     421 KLLIKKNSATEIEEENLCDSLLKDQEAAYKGQNKCVKVDNLFWFQCADGYTTTYEMT 477


RESULT 15
US-08-461-004A-56
; Sequence 56, Application US/08461004A
; Patent No. 6235283
;  GENERAL INFORMATION:
;    APPLICANT:  COBON, Stewart Gary
;    APPLICANT:  MOORE, Joanna Terry
;    APPLICANT:  JOHNSON, Law Anthony York
;    APPLICANT:  WILLADSEN, Peter
;    APPLICANT:  KEMP, David Harold
;    APPLICANT:  SRISKANTHA, Alagacone
;    APPLICANT:  RIDING, George Alfred
;    APPLICANT:  RAND, Keith No. 6235283man
;    TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;    TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,004A
;      FILING DATE:  04-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/325,071
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,109
;      FILING DATE:  17-MAY-1993
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US 07/926,368
;       FILING DATE:  07-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/242,196
;       FILING DATE:  06-JUL-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/AU87/00401
;       FILING DATE:  27-NOV-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI4912
;       FILING DATE:  16-OCT-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PI2570
;       FILING DATE:  19-JUN-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PH9196
;       FILING DATE:  27-NOV-1986
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  60042/152
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 672 5300
;       TELEFAX:  202 672 5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  56:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  650 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-461-004A-56

  Query Match             7.0%;  Score 70;  DB 4;  Length 650;
  Best Local Similarity   22.6%;  Pred. No. 8.9;
  Matches   40;  Conservative   28;  Mismatches   51;  Indels   58;  Gaps   11;

Qy      17 ELLDCGYHPESLLSDFDYWDYVVPEPN-------LNEVIFEESTCQNL------VKMLEN 63
           | :||| :  :  | :  |    |/||       |||  :  |   |:       |  |:
Db     303 EFMDCGVY-MNRQSCYCPWKSRKPGPNVVNINECLLNEYYYTVSFTPNISFDSDHCKWYED 361

Qy      64 CLSKSKQTKLG----------CSKVLVPEKLTQR-IAQDVLR-LSSTE-PCGLRGCVMHV 110
             : :: :| :|          |:: :    : :: ::: ||| | : | | |   |:|:
Db     362 RVLEAIRTSIGKEVFKVEILNCTQDIKARLIAEKPLSKHVLRKLQACEHPIG-EWCMMYP 420

Qy     111 NL--------EIE------------------NVCKKLDRIV---CDSSVVPTFELT 137
           |        |||                  | | |:| :    |      |:|:|
Db     421 KLLIKKNSATEIEEENLCDSLLKDQEAAYKGQNKCVKVDNLFWFQCADGYTTTYEMT 477


Search completed: October 11, 2002, 14:15:33
Job time : 34 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 11, 2002, 12:55:07 ; Search time 35 Seconds
(without alignments)
213.511 Million cell updates/sec

Title: US-09-881-636-2
Perfect score: 1002
Sequence: 1 MVATGSLSSKNPASISELLD.....FRLVKKKLYSLIGTTVIEGS 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90.5	9.0	1056	1	RIC1_YEAST	P40395 saccharomyc
2	85	8.5	1098	1	RP16_DROME	Q9gqn0 drosophila
3	80	8.0	1416	1	BLM_MOUSE	O88700 mus musculu
4	79	7.9	315	1	CCG3_HUMAN	O60359 homo sapien
5	79	7.9	549	1	CED4_CAEEL	P30429 caenorhabdi
6	78.5	7.8	463	1	GLGA_AQUAE	O66935 aquifex aeo
7	78.5	7.8	964	1	LON2_MAIZE	P93648 zea mays (m
8	77.5	7.7	356	1	Y4ZB_RHISN	P55729 rhizobium s
9	77	7.7	1417	1	BLM_HUMAN	P54132 homo sapien
10	77	7.7	1828	1	MY5A_RAT	Q9qyf3 rattus norv
11	77	7.7	1829	1	MY5A_CHICK	Q02440 gallus gall
12	77	7.7	1853	1	MY5A_MOUSE	Q99104 mus musculu
13	77	7.7	1855	1	MY5A_HUMAN	Q9y4i1 homo sapien
14	76	7.6	1390	1	RPOB_MYCGA	P47715 mycoplasma
15	75.5	7.5	613	1	GP37_HUMAN	O15354 homo sapien
16	75	7.5	418	1	FTSA_BUCAP	O51928 buchnera ap
17	75	7.5	439	1	VP51_BPAPS	Q9t1p7 bacteriopha
18	74.5	7.4	323	1	CCG2_HUMAN	Q9y698 homo sapien
19	74.5	7.4	323	1	CCG2_MOUSE	O88602 mus musculu
20	74.5	7.4	676	1	TIM_DROHY	O44431 drosophila
21	74.5	7.4	4196	1	DYHC_SCHPO	O13290 schizosacch
22	73.5	7.3	366	1	FLHF_BACSU	Q01960 bacillus su
23	73.5	7.3	441	1	YD40_HUMAN	Q9p2k6 homo sapien
24	73	7.3	145	1	RL15_BORBU	O51450 borrelia bu
25	73	7.3	334	1	DCAM_BOVIN	P50243 bos taurus
26	73	7.3	334	1	DCAM_MESAU	P28918 mesocricetu
27	73	7.3	664	1	SL51_SHEEP	P53791 ovis aries
28	72.5	7.2	610	1	VE1_HPV60	Q80943 human papil
29	72	7.2	334	1	DCAM_HUMAN	P17707 homo sapien
30	72	7.2	486	1	VN53_ROTHW	P35424 human rotav
31	72	7.2	530	1	UDB2_RAT	P08541 rattus norv
32	72	7.2	1640	1	CLH2_HUMAN	P53675 homo sapien
33	71.5	7.1	410	1	Y474_AQUAE	O66772 aquifex aeo
34	71.5	7.1	510	1	GARD_BACSU	P42240 bacillus su
35	71.5	7.1	566	1	TS13_MOUSE	Q01755 mus musculu
36	71.5	7.1	1818	1	Z294_HUMAN	O94822 homo sapien
37	71	7.1	179	1	RL5_BUCAK	P46178 buchnera ap
38	71	7.1	469	1	VL2_BPV1	P03109 bovine papi
39	71	7.1	814	1	SYV_RICPR	Q9zcn6 rickettsia
40	71	7.1	1005	1	E4L2_HUMAN	O43491 homo sapien
41	70.5	7.0	333	1	DCAM_RAT	P17708 rattus norv
42	70.5	7.0	334	1	DCM1_MOUSE	P31154 mus musculu
43	70.5	7.0	334	1	DCM2_MOUSE	P82184 mus musculu
44	70.5	7.0	334	1	DCM2_MUSSP	P82185 mus spretus
45	70.5	7.0	530	1	UDBH_HUMAN	O75795 homo sapien

ALIGNMENTS

RESULT 1
RIC1_YEAST
ID RIC1_YEAST STANDARD; PRT; 1056 AA.
AC P40395;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RIC1 protein.
GN RIC1 OR YLR039C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mizuta K.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Koetter P., Rose M., Entian K.D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTION OF BOTH RIBOSOMAL PROTEIN
CC GENES AND RIBOSOMAL RNA.
CC ---

CC ---
DR EMBL; D43895; BAA07866.1; -.
DR EMBL; Z73211; CAA97567.1; -.
DR TRANSFAC; T03560; -.
DR SGD; S0004029; RIC1.
KW Transcription regulation.
SQ SEQUENCE 1056 AA; 121649 MW; 2368616B67AEAD31 CRC64;

Query Match 9.0%; Score 90.5; DB 1; Length 1056;
Best Local Similarity 25.0%; Pred. No. 1.3;
Matches 42; Conservative 23; Mismatches 52; Indels 51; Gaps 8;

```
Qy      24 HPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEK 83
           :|:  |||:  |:| |||   | ::|          :  | ||: |  |:        ||
Db     613 NPDVELSDYVMWEYNVPE---NTIVF---------SLHVNTLSRYKLLKMKSKNHNASEK 660

Qy      84 L---------------TQRIAQDVLRLSSTEPCGLRGC-----VMHVNLEIENVCKKLDR 123
                          || |  ||  :|:  ||||         : :|: |: :  |::
Db     661 QPDALLKTAEIILVTDTQTIVFDV--ISTVHPCGLNIIKKFYQYLKINIPIDVLPNKIEW 718

Qy     124 IVCDSSVVPTFELTLVFKQENCSWTSFRD-FFFSRGRFSSGFRRTLIL 170
           |:        :  |:|         | |  |   |:   |  :|| |
Db     719 II-------NMKEGLLF---------FADRKFIKLGKVDGGGWQTLTL 750
```

```
RESULT 2
RP16_DROME
ID   RP16_DROME     STANDARD;      PRT;  1098 AA.
AC   Q9GQN0; Q9VXL6;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ran-binding protein 16.
GN   RANBP16 OR CG9126.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goerlich D., Hartmann E.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: May function as a nuclear transport receptor (By
CC       similarity).
CC   -!- SUBUNIT: Binds to nucleoporins and the GTP-bound form of Ran (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, nuclear and nuclear pore
CC       complex (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE EXPORTIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF222744; AAG44254.1; -.
DR   EMBL; AE003500; AAF48542.1; ALT_SEQ.
DR   FlyBase; FBgn0030709; Ranbp16.
KW   Nuclear protein; Transport; Protein transport.
SQ   SEQUENCE   1098 AA;  125532 MW;  C8A3885CB3FDAEF5 CRC64;

  Query Match             8.5%;  Score 85;  DB 1;  Length 1098;
  Best Local Similarity   25.6%;  Pred. No. 4.5;
  Matches   30;  Conservative   21;  Mismatches   44;  Indels   22;  Gaps    4;

Qy      36 DYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRL 95
           :|:   |||   :      | || :|      :| || |       | : | : :|| :
Db      83 NYLATVPNLQHFV-----VQALVSLL------AKLTKYGWFDSYKEEMVFQNLLEDVKKF 131

Qy      96 SSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRD 152
                ||:| | |  : ::     | ::||: : |   :: : | :     |||||
Db     132 -------LQGSVEHCTIGVQ----ILSQLVCEMNSVVEMDVQVSFSKMRKIATSFRD 177


RESULT 3
BLM_MOUSE
ID   BLM_MOUSE      STANDARD;      PRT;  1416 AA.
AC   O88700; O88198;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bloom's syndrome protein homolog (EC 3.6.1.-) (mBLM).
GN   BLM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS.
RX   MEDLINE=99054654; PubMed=9840919;
RA   Bahr A., de Graeve F., Kedinger C., Chatton B.;
RT   "Point mutations causing Bloom's syndrome abolish ATPase and DNA
RT   helicase activities of the BLM protein.";
RL   Oncogene 17:2565-2571(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=BALB/C; TISSUE=Testis, Spermatocyte, and Brain;
RX   MEDLINE=98322127; PubMed=9655940;
RA   Seki T., Wang W.-S., Okumura N., Seki M., Katada T., Enomoto T.;
RT   "cDNA cloning of mouse BLM gene, the homologue to human Bloom's
RT   syndrome gene, which is highly expressed in the testis at the mRNA
RT   level.";
RL   Biochim. Biophys. Acta 1398:377-381(1998).
CC   -!- FUNCTION: PARTICIPATES IN DNA REPLICATION AND MAY PARTICIPATE IN
CC       REPAIR. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-HELICASE
CC       ACTIVITY THAT UNWINDS SINGLE- AND DOUBLE-STRANDED DNA IN A 3'-5'
CC       DIRECTION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS 12-14 DAYS AFTER
CC       BIRTH (CORRESPONDING TO THE PACHYTENE PHASE) AND AT MUCH LOWER
CC       LEVELS IN BRAIN, HEART, KIVER, LUNG, THYMUS, KIDNEY AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HDRC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z98263; CAB10933.1; -.
DR   EMBL; AB008674; BAA32001.1; -.
DR   MGD; MGI:1328362; Blm.
```

```
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_ATP_helicase.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW   Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW   DNA replication.
FT   DOMAIN      219    231       POLY-GLU.
FT   DOMAIN      564    574       POLY-ASP.
FT   DOMAIN     1312   1317       POLY-GLU.
FT   NP_BIND     697    704       ATP (BY SIMILARITY).
FT   SITE        803    806       DEAH BOX.
FT   DOMAIN     1217   1297       HDRC.
FT   DOMAIN     1333   1348       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   MUTAGEN     680    680       Q->P: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN     703    703       K->A: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN     849    849       I->T: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   MUTAGEN    1063   1063       C->S: REDUCED ATPASE AND HELICASE
FT                                ACTIVITIES.
FT   CONFLICT    131    131       L -> P (IN REF. 2).
FT   CONFLICT    229    229       E -> EE (IN REF. 2).
FT   CONFLICT    535    535       V -> M (IN REF. 2).
FT   CONFLICT    546    547       WN -> RT (IN REF. 2).
FT   CONFLICT    574    574       MISSING (IN REF. 2).
FT   CONFLICT    591    591       T -> A (IN REF. 2).
FT   CONFLICT    621    621       T -> N (IN REF. 2).
FT   CONFLICT   1295   1295       V -> L (IN REF. 2).
SQ   SEQUENCE   1416 AA;  158365 MW;  447C8110A775DD42 CRC64;

  Query Match             8.0%;  Score 80;  DB 1;  Length 1416;
  Best Local Similarity   24.5%;  Pred. No. 18;
  Matches   38;  Conservative   32;  Mismatches   43;  Indels   42;  Gaps   11;

Qy      30 SDFDYWDYVVPEPNLNEVIFEEST---CQNLVKMLENCLSKSKQTKLGCSKVLV--PEKL 84
           :|:|  |:| | |   |:|   |:   | :::| |::  |  ::  |  |: |:  ||::
Db     299 NDYDI-DFVPPSP--EEIISTASSSLKCSSMLKDLDD--SDKEKGILSTSEELLSKPEEM 353

Qy      85 TQRIAQDVLRLSSTEPCG-----LRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLV 139
           |   :       ::: |      ::  ::||   :|::|| :|        ||| ||  :
Db     354 TTHKSD----AGTSKDCDAQQIRIQQQLIHV---MEHICKLVD--------TVPTDELEAL 399

Qy     140 FKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGF 174
             ||    :           || |:  :||
Db     400 ----NCGTELLQ---------QRNIRRKLLAEAGF 421


RESULT 4
CCG3_HUMAN
ID   CCG3_HUMAN     STANDARD;      PRT;   315 AA.
AC   O60359;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Voltage-dependent calcium channel gamma-3 subunit (Neuronal voltage-
DE   gated calcium channel gamma-3 subunit).
GN   CACNG3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RX   MEDLINE=99236653; PubMed=10221464;
RA   Black J.L. III, Lennon V.A.;
RT   "Identification and cloning of putative human neuronal voltage-gated
RT   calcium channel gamma-2 and gamma-3 subunits: neurologic
RT   implications.";
RL   Mayo Clin. Proc. 74:357-361(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99425270; PubMed=10493829;
RA   Loftus B.J., Kim U.J., Sneddon V.P., Kalush F., Brandon R.,
RA   Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA   Deslattes Mays A., Cao Y., Xu R.X., Kang H.L., Mitchell S.,
RA   Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT   "Genome duplications and other features in 12 Mb of DNA sequence from
RT   human chromosome 16p and 16q.";
RL   Genomics 60:295-308(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain cortex;
RA   Xia J.-H., Zhang H.-L., Tang X.-X., Yu K.-P., Pan Q., Dai H.-P.;
RT   "Cloning of human calcium channel gamma-3 subunit.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THOUGHT TO STABILIZE THE CALCIUM CHANNEL IN AN
CC       INACTIVATED (CLOSED) STATE.
CC   -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FIVE SUBUNITS:
CC       ALPHA-1, ALPHA-2/DELTA, BETA AND GAMMA.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DISEASE: CANDIDATE GENE FOR A FAMILIAL INFANTILE CONVULSIVE
CC       DISORDER WITH PAROXYSOMAL CHOREOATHETOSIS.
CC   -!- SIMILARITY: BELONGS TO THE PMP-22 / EMP / MP20 FAMILY. CACNG
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF100346; AAD22739.1; -.
DR   EMBL; AC004125; AAC15246.1; -.
DR   EMBL; AF134640; AAF42975.1; -.
DR   MIM; 606403; -.
DR   InterPro; IPR004031; PMP22_Claudin.
DR   Pfam; PF00822; PMP22_Claudin; 1.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM    104    124       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
SQ   SEQUENCE   315 AA;  35548 MW;  5DCFB9FB44B5A2F4 CRC64;

  Query Match             7.9%;  Score 79;  DB 1;  Length 315;
  Best Local Similarity   23.4%;  Pred. No. 3.6;
  Matches   41;  Conservative   16;  Mismatches   50;  Indels   68;  Gaps    8;

Qy      33 DYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDV 92
           |||            ::    |            ::| |          :  | |  ::|
Db      31 DYW-----------LYSRGVC-----------RTKSTS---------DNETSRKNEEV 57

Qy      93 LRLSST-EPCGLRGCVMHVNLEIENVCKKLD----------------RIVCDSSVVPTF 134
           :  |     | | |           ||||:|                  | |  ||| |
Db      58 MTHSGLWRTCCLEGA-------FRGVCKKIDHFPEDADYEQDTAEYLLRAVRASSVFPIL 110

Qy     135 ELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIGTTV 189
            :||:|    |   |  :|  |        |  :|||:|   |   | ::||  |
Db     111 SVTLLFFGGLCVAAS--EFHRS--------RHNVILSAGIFFVSAGLSNIIGIIV 155


RESULT 5
CED4_CAEEL
ID   CED4_CAEEL     STANDARD;      PRT;   549 AA.
```

```
AC   P30429;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cell death protein 4.
GN   CED-4 OR C35D10.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93161939; PubMed=1286611;
RA   Yuan J., Horvitz H.R.;
RT   "The Caenorhabditis elegans cell death gene ced-4 encodes a novel
RT   protein and is expressed during the period of extensive programmed
RT   cell death.";
RL   Development 116:309-320(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton L.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXPRESSED DURING THE PERIOD OF EXTENSIVE PROGRAMMED
CC       CELL DEATH. CED-4, ALONG WITH CED-3, SEEMS TO FUNCTION WITHIN
CC       DYING CELLS TO CAUSE CELL DEATH.
CC   -!- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING EMBRYOGENESIS AND IS
CC       ALSO DETECTED AT LATER STAGES.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69016; CAA48781.1; -.
DR   EMBL; U21324; AAA62564.1; -.
DR   WormPep; C35D10.9; CE01203.
DR   InterPro; IPR001315; CARD.
DR   InterPro; IPR002182; NB-ARC.
DR   Pfam; PF00619; CARD; 1.
DR   Pfam; PF00931; NB-ARC; 1.
DR   SMART; SM00114; CARD; 1.
DR   PROSITE; PS50209; CARD; 1.
KW   Apoptosis; ATP-binding.
FT   DOMAIN       1     91       CARD.
FT   NP_BIND     159    166      ATP (POTENTIAL).
SQ   SEQUENCE   549 AA;  62878 MW;  DB2A7969BDA50AF8 CRC64;

  Query Match             7.9%;  Score 79;  DB 1;  Length 549;
  Best Local Similarity   24.7%;  Pred. No. 7;
  Matches   45;  Conservative   23;  Mismatches   46;  Indels   68;  Gaps   11;

Qy      10 KNPASISELLD-CGYHPESLLSDF--DYWDYVVPEPNLNE--VIFEESTCQNLVKMLENC 64
           :  : :  |:|   |: :| |:||  || |: : ||:|    ||  : : | |
Db      63 RQASELGPLIDFFNYNNQSHLADFLEDYIDFAINEPDLLRPVVIAPQFSRQML------- 115

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
             ::  ||     ||:::|                | :|         :: | ||||
Db     116 ---DRKLLLG----NVPKQMT--------------CYIR------EYHVDRVIKKLD-- 145

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSL 184
                       | |   ||  | |  ||  ||  :::| |         |   |
Db     146 ------------------EMCDLDSF--FLFLHGRAGSG--KSVIASQAL----SKSDQL 179

Qy     185 IG 186
           ||
Db     180 IG 181
```

```
RESULT 6
GLGA_AQUAE
ID   GLGA_AQUAE     STANDARD;      PRT;   463 AA.
AC   O66935;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Glycogen synthase (EC 2.4.1.21) (Starch [bacterial glycogen]
DE   synthase).
GN   GLGA OR AQ_721.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: Synthesizes alpha-1,4-glucan chains using ADP-glucose.
CC   -!- CATALYTIC ACTIVITY: ADP-glucose + [(1,4)-alpha-D-glucosyl](N) =
CC       ADP + [(1,4)-alpha-D-glucosyl](N+1).
CC   -!- PATHWAY: Glycogen biosynthesis; second step.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000704; AAC06894.1; -.
DR   InterPro; IPR001296; Glycos_transf_1.
DR   Pfam; PF00534; Glycos_transf_1; 1.
KW   Glycogen biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.
FT   BINDING      15     15       ADP-GLUCOSE (BY SIMILARITY).
SQ   SEQUENCE   463 AA;  53457 MW;  D896172D0D3DB27D CRC64;

  Query Match             7.8%;  Score 78.5;  DB 1;  Length 463;
  Best Local Similarity   19.6%;  Pred. No. 6.4;
  Matches   37;  Conservative   37;  Mismatches   78;  Indels   37;  Gaps    5;

Qy       1 MVATGSLSSKNPASISEL--LDCGYHPESLLSDFDY--------WDYVVPEPNLNEVIFE 50
             :| :  ::: :|    |:   : ||  | :|  : |          || |  |  :: |::
Db     198 IVFSDLITTVSPTYAKEIQTQEYGYGLEGVLKKYSYKLRGILNGIDYEVWNPEKDKYIYQ 257

Qy      51 ESTCQNLVKMLENCLSKSKQTKLGCSKVLVP--EKLTQRIAQDVLRLSSTEPCGLRGCVM 108
             : :|  |  :|    ||:  :   | |:    : | :   ::
Db     258 NYSLRNYSKKFKNKEFLSKELGIEAEKPLISFINRFTHQKGVEL---------------- 301

Qy     109 HVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTL 168
               | |  ::: ::  :   : | |    |   :  : :|: |      |: || | |
Db     302 -----ILNCAEEMSKLNANFVFLGTGEYENAFLDVSKIYKNFKVF----AEFNEGFARKL 352

Qy     169 ILSSGFRLV 177
             || | |:
Db     353 YASSDFILM 361


RESULT 7
LON2_MAIZE
ID   LON2_MAIZE     STANDARD;      PRT;   964 AA.
AC   P93648;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
```

```
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lon protease homolog 2, mitochondrial precursor (EC 3.4.21.-).
GN   LON2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. B73;
RA   Rapp W.D., Barakat S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LON
CC       FAMILY OF ATP-DEPENDENT PROTEASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U85495; AAC50021.1; -.
DR   MEROPS; S16.002; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003959; AAA_subfam.
DR   InterPro; IPR003111; LON.
DR   InterPro; IPR001984; Lon_endopep.
DR   Pfam; PF00004; AAA; 1.
DR   Pfam; PF02190; LON; 1.
DR   PRINTS; PR00830; ENDOLAPTASE.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00464; LON; 1.
DR   PROSITE; PS01046; LON_SER; 1.
KW   Hydrolase; Serine protease; ATP-binding; Multigene family;
KW   Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    964       LON PROTEASE HOMOLOG 2.
FT   NP_BIND     455    462       ATP (POTENTIAL).
FT   ACT_SITE    863    863       BY SIMILARITY.
SQ   SEQUENCE   964 AA;  105659 MW;  D78A6C6B0F8A6D9E CRC64;

  Query Match             7.8%;  Score 78.5;  DB 1;  Length 964;
  Best Local Similarity   27.6%;  Pred. No. 15;
  Matches   34;  Conservative   28;  Mismatches   42;  Indels   19;  Gaps    8;

Qy       5 GSLSSKNPAS-ISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTC-QNLVKMLE 62
           |   | :||| : |||| |  ||  ::  |:  |:    :|::|:|   | |:::|:
Db     528 GKGHSGDPASALLELLD----PEQNVNFLDH--YLDVPIDLSKVLF---VCTANVIEMIP 578

Qy      63 NCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVN-----LEIENV 117
           | |   :   :  :  : ||:   ||:| |  :: : ||::   : |       |||
Db     579 NPL-LDRMEIIAIAGYITDEKM--HIARDYLEKNTRQACGIKPEQVEVTDTALLALIENY 635

Qy     118 CKK 120
           |::
Db     636 CRE 638


RESULT 8
Y4ZB_RHISN
ID   Y4ZB_RHISN     STANDARD;      PRT;   356 AA.
AC   P55729;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative transposase Y4ZB.
GN   Y4ZB.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 11.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000108; AAB91960.1; -.
DR   InterPro; IPR002559; Transposase_11.
DR   Pfam; PF01609; Transposase_11; 1.
KW   Hypothetical protein; Transposable element; Transposition;
KW   DNA-binding; DNA recombination; Plasmid.
SQ   SEQUENCE   356 AA;  39775 MW;  9F350ABB7E691635 CRC64;

  Query Match             7.7%;  Score 77.5;  DB 1;  Length 356;
  Best Local Similarity   25.2%;  Pred. No. 5.7;
  Matches   28;  Conservative   22;  Mismatches   34;  Indels   27;  Gaps    6;

Qy      82 EKLTQRIAQDVLRLSSTEP------CG-------LRGCVMHVNLEIENVCKKL----DRI 124
           :: |:|  : :|||  : |       ||       :||  |||  : :: | :|    |
Db      76 DRQTRREGRAMLRLIDSTPIPLGKLCGWAKSNGRIRGMKMHVVYDPDSDCPRLLDITDAN 135

Qy     125 VCDSSVVPTFEL----TLVFKQENCS---WTSFRD---FFFSRGRFSSGFR 165
           | |: :  |  :    | :| :  |    ||:  :   || :| : : | :
Db     136 VNDAQIGRTIAIESGATYIFDKGYCHYGWWTAIAEAKAFFVTRPKSNMGLK 186


RESULT 9
BLM_HUMAN
ID   BLM_HUMAN      STANDARD;      PRT;  1417 AA.
AC   P54132;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bloom's syndrome protein (EC 3.6.1.-) (RecQ protein-like 3) (DNA
DE   helicase, RecQ-like, type 2).
GN   BLM OR RECQL3 OR RECQ2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS BS ARG-672; ILE-843 AND SER-1055.
RX   MEDLINE=96069866; PubMed=7585968;
RA   Ellis N.A., Groden J., Ye T.-Z., Straughen J., Lennon D.J., Ciocci S.,
RA   Proytcheva M., German J.;
RT   "The Bloom's syndrome gene product is homologous to RecQ helicases.";
RL   Cell 83:655-666(1995).
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RC   TISSUE=B-cell;
RX   MEDLINE=98049515; PubMed=9388193;
RA   Karow J.K., Chakraverty R.K., Hickson I.D.;
RT   "The Bloom's syndrome gene product is a 3'-5' DNA helicase.";
RL   J. Biol. Chem. 272:30611-30614(1997).
RN   [3]
RP   NUCLEAR LOCALIZATION SIGNAL.
RX   MEDLINE=98049834; PubMed=9388480;
RA   Kaneko H., Orii K.O., Matsui E., Shimozawa N., Fukao T., Matsumoto T.,
RA   Shimamoto A., Furuichi Y., Hayakawa S., Kasahara K., Kondo N.;
```

```
RT   "BLM (the causative gene of Bloom syndrome) protein translocation into
RT   the nucleus by a nuclear localization signal.";
RL   Biochem. Biophys. Res. Commun. 240:348-353(1997).
RN   [4]
RP   VARIANT BS PHE-1036.
RX   MEDLINE=97449163; PubMed=9285778;
RA   Foucault F., Vaury C., Barakat A., Thibout D., Planchon P., Jaulin C.,
RA   Praz F., Amor-Gueret M.;
RT   "Characterization of a new BLM mutation associated with a
RT   topoisomerase II alpha defect in a patient with Bloom's syndrome.";
RL   Hum. Mol. Genet. 6:1427-1434(1997).
CC   -!- FUNCTION: PARTICIPATES IN DNA REPLICATION AND MAY PARTICIPATE IN
CC       REPAIR. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-HELICASE
CC       ACTIVITY THAT UNWINDS SINGLE- AND DOUBLE-STRANDED DNA IN A 3'-5'
CC       DIRECTION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: DEFECTS IN BLM ARE THE CAUSE OF BLOOM SYNDROME (BS), AN
CC       AUTOSOMAL RECESSIVE DISORDER CHARACTERIZED BY PROPORTIONATE PRE-
CC       AND POSTNATAL GROWTH DEFICIENCY; SUN-SENSITIVE, TELANGIECTATIC,
CC       HYPO- AND HYPERPIGMENTED SKIN; PREDISPOSITION TO MALIGNANCY; AND
CC       CHROMOSOMAL INSTABILITY.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HDRC DOMAIN.
CC   -!- DATABASE: NAME=BLMbase; NOTE=BLM mutation db;
CC       WWW="http://www.uta.fi/imt/bioinfo/BLMbase/".
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U39817; AAA87850.1; -.
DR   MIM; 604610; -.
DR   MIM; 210900; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_ATP_helicase.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW   Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW   DNA replication; Disease mutation.
FT   DOMAIN      292    299       POLY-ASP.
FT   DOMAIN      310    316       POLY-SER.
FT   DOMAIN      557    566       POLY-ASP.
FT   NP_BIND     689    696       ATP (BY SIMILARITY).
FT   SITE        795    798       DEAH BOX.
FT   DOMAIN     1212   1292       HDRC.
FT   DOMAIN     1334   1349       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARIANT     672    672       Q -> R (IN BS).
FT                                /FTId=VAR_006901.
FT   VARIANT     843    843       T -> I (IN BS).
FT                                /FTId=VAR_006902.
FT   VARIANT     891    891       G -> E (IN BS).
FT                                /FTId=VAR_009138.
FT   VARIANT     901    901       C -> Y (IN BS).
FT                                /FTId=VAR_009139.
FT   VARIANT    1036   1036       C -> F (IN BS).
FT                                /FTId=VAR_009140.
FT   VARIANT    1055   1055       C -> S (IN BS).
FT                                /FTId=VAR_006903.
SQ   SEQUENCE   1417 AA;  158999 MW;  423DF5F381194E11 CRC64;
```

- Query Match 7.7%; Score 77; DB 1; Length 1417;
  Best Local Similarity 31.8%; Pred. No. 34;

```
  Matches   34;  Conservative   21;  Mismatches   32;  Indels   20;  Gaps    8;

Qy      31 DFDY-WDYVVPEPNLNEVIFEEST---CQNLVKMLENCLSKSKQTKLGCSKVLV--PEKL 84
           | ||  |:| | |   |:|   |:   | : :| |:   |  |:  |  || |:  |||:
Db     293 DDDYDTDFVPPSP--EEIISASSSSSKCLSTLKDLDT--SDRKEDVLSTSKDLLSKPEKM 348

Qy      85 TQRIAQDVLRLSSTE----PCGLRGCVMHVNLEIENVCKKLDRIVCD 127
           :    |::   :||:       |:  ::||   :|::|| :| |  |
Db     349 SM---QELNPETSTDCDARQISLQQQLIHV---MEHICKLIDTIPDD 389


RESULT 10
MY5A_RAT
ID   MY5A_RAT       STANDARD;      PRT;  1828 AA.
AC   Q9QYF3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
GN   MYO5A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Futaki S., Murata Y., Hayashi Y.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC       STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC       MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC       BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC       FORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC       CALMODULIN OR MYOSIN LIGHT CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AB035736; BAA88350.1; -.
DR   HSSP; P08799; 1LVK.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00015; IQ; 6.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 6.
KW   Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Phosphorylation.
FT   DOMAIN        1    765       MYOSIN HEAD-LIKE.
FT   DOMAIN      766    788       IQ 1.
FT   DOMAIN      789    813       IQ 2.
FT   DOMAIN      814    836       IQ 3.
FT   DOMAIN      837    861       IQ 4.
FT   DOMAIN      862    884       IQ 5.
FT   DOMAIN      885    914       IQ 6.
FT   DOMAIN      914   1239       COILED COIL (POTENTIAL).
FT   DOMAIN     1314   1418       COILED COIL (POTENTIAL).
FT   DOMAIN     1660   1765       DILUTE.
```

```
FT   NP_BIND     163    170       ATP (POTENTIAL).
FT   DOMAIN      643    665       ACTIN-BINDING (POTENTIAL).
FT   MOD_RES    1733   1733       PHOSPHORYLATION (POTENTIAL).
SQ   SEQUENCE   1828 AA;  211762 MW;  5B3DE1C89AE36123 CRC64;

  Query Match             7.7%;  Score 77;  DB 1;  Length 1828;
  Best Local Similarity   21.8%;  Pred. No. 46;
  Matches   42;  Conservative   33;  Mismatches   60;  Indels   58;  Gaps    9;

Qy      26 ESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLT 85
           |  |::||  :|     :|   |:     | ||::||| |           ::|   |
Db    1560 EHCLTNFDLAEYRQVLSDLAIQIY-----QQLVRVLENIL----------QPMIVSGMLE 1604

Qy      86 QRIAQDVLRLSSTEPCGLRGCVMHVNLE----IENVCKKLD---RIVCDSSVVPTFEL-- 136
             | |   |  :| |||     :  |    :::: ::|:    ::|   : |  ||
Db    1605 HETIQGV---SGVKPTGLRKRTSSIADEGTYTLDSILRQLNSFHSVMCQHGMDP--ELIK 1659

Qy     137 ---------------TLVFKQENCSW----------TSFRDFFFSRGRFSSGFRRT--- 167
                           |: ::: |||          :   ::   :   :|| : |
Db    1660 QVVKQMFYIVGAITLNNLLLRKDMCSWSKGMQIRYNVSQLEEWLRDKNLMNSGAKETLEP 1719

Qy     168 LILSSGFRLVKKK 180
           || ::    ||||
Db    1720 LIQAAQLLQVKKK 1732


RESULT 11
MY5A_CHICK
ID   MY5A_CHICK     STANDARD;      PRT;  1829 AA.
AC   Q02440;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
DE   heavy chain P190) (Myosin-V).
GN   MYO5A.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93012002; PubMed=1383040;
RA   Sanders G., Lichte B., Meyer H.E., Kilimann M.W.;
RT   "cDNA encoding the chicken ortholog of the mouse dilute gene product.
RT   Sequence comparison reveals a myosin I subfamily with conserved C-
RT   terminal domains.";
RL   FEBS Lett. 311:295-298(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC   TISSUE=Brain;
RX   MEDLINE=93107155; PubMed=1469047;
RA   Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA   de Camilli P.V., Larson R.E., Mooseker M.S.;
RT   "Primary structure and cellular localization of chicken brain
RT   myosin-V (p190), an unconventional myosin with calmodulin light
RT   chains.";
RL   J. Cell Biol. 119:1541-1557(1992).
CC   -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC       STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC       MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC       BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC       FORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC       CALMODULIN OR MYOSIN LIGHT CHAINS.
CC   -!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC       (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67251; CAA47673.1; -.
DR   EMBL; Z11718; CAA77782.1; -.
DR   HSSP; P08799; 1MMG.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00015; IQ; 6.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 6.
KW   Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Phosphorylation.
FT   DOMAIN        1    766       MYOSIN HEAD-LIKE.
FT   DOMAIN      767    789       IQ 1.
FT   DOMAIN      790    814       IQ 2.
FT   DOMAIN      815    837       IQ 3.
FT   DOMAIN      838    862       IQ 4.
FT   DOMAIN      863    887       IQ 5.
FT   DOMAIN      888    915       IQ 6.
FT   DOMAIN      916   1239       COILED COIL (POTENTIAL).
FT   DOMAIN     1315   1419       COILED COIL (POTENTIAL).
FT   DOMAIN     1661   1766       DILUTE.
FT   NP_BIND     163    170       ATP (BY SIMILARITY).
FT   DOMAIN      644    666       ACTIN-BINDING (POTENTIAL).
FT   MOD_RES    1734   1734       PHOSPHORYLATION (POTENTIAL).
FT   CONFLICT   1142   1142       E -> EQ (IN REF. 2).
SQ   SEQUENCE   1829 AA;  212381 MW;  0538B278DFC09F6E CRC64;

  Query Match             7.7%;  Score 77;  DB 1;  Length 1829;
  Best Local Similarity   21.8%;  Pred. No. 46;
  Matches   42;  Conservative   33;  Mismatches   60;  Indels   58;  Gaps    9;

Qy      26 ESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLT 85
           |  |::||  :|     :|   |:     | ||::||| |           ::|   |
Db    1561 EHCLTNFDLAEYRQVLSDLAIQIY-----QQLVRVLENIL----------QPMIVSGMLE 1605

Qy      86 QRIAQDVLRLSSTEPCGLRGCVMHVNLE----IENVCKKLD---RIVCDSSVVPTFEL-- 136
             | |   |  :| |||     :  |    :::: ::|:    ::|   : |  ||
Db    1606 HETIQGV---SGVKPTGLRKRTSSIADEGTYTLDSIIRQLNSFHSVMCQHGMDP--ELIK 1660

Qy     137 ---------------TLVFKQENCSW----------TSFRDFFFSRGRFSSGFRRT--- 167
                           |: ::: |||          :   ::   :   :|| : |
Db    1661 QVVKQMFYIIGAVTLNNLLLRKDMCSWSKGMQIRYNVSQLEEWLRDKNLMNSGAKETLEP 1720

Qy     168 LILSSGFRLVKKK 180
           || ::    ||||
Db    1721 LIQAAQLLQVKKK 1733


RESULT 12
MY5A_MOUSE
ID   MY5A_MOUSE     STANDARD;      PRT;  1853 AA.
AC   Q99104;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
GN   MYO5A OR DILUTE.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=91141583; PubMed=1996138;
RA   Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA   Jenkins N.A.;
RT   "Novel myosin heavy chain encoded by murine dilute coat colour
RT   locus.";
RL   Nature 349:709-712(1991).
RN   [2]
RP   REVISIONS.
RA   Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA   Jenkins N.A.;
RL   Nature 352:547-547(1991).
CC   -!- FUNCTION: MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR
CC       ALTERNATIVELY, IT MAY BE REQUIRED FOR SOME POLARIZATION PROCESS
CC       INVOLVED IN DENDRITE FORMATION.
CC   -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC       CALMODULIN OR MYOSIN LIGHT CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57377; CAA40651.1; -.
DR   PIR; A46761; A46761.
DR   HSSP; P08799; 1LVK.
DR   MGD; MGI:105976; Myo5a.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00015; IQ; 6.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 6.
KW   Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Phosphorylation.
FT   DOMAIN        1    765       MYOSIN HEAD-LIKE.
FT   DOMAIN      766    788       IQ 1.
FT   DOMAIN      789    813       IQ 2.
FT   DOMAIN      814    836       IQ 3.
FT   DOMAIN      837    861       IQ 4.
FT   DOMAIN      862    884       IQ 5.
FT   DOMAIN      885    913       IQ 6.
FT   DOMAIN      914   1237       COILED COIL (POTENTIAL).
FT   DOMAIN     1314   1443       COILED COIL (POTENTIAL).
FT   DOMAIN     1685   1790       DILUTE.
FT   NP_BIND     163    170       ATP (POTENTIAL).
FT   DOMAIN      643    665       ACTIN-BINDING (POTENTIAL).
FT   MOD_RES    1758   1758       PHOSPHORYLATION (POTENTIAL).
SQ   SEQUENCE   1853 AA;  215594 MW;  503E93D48CA6B766 CRC64;

  Query Match             7.7%;  Score 77;  DB 1;  Length 1853;
  Best Local Similarity   21.8%;  Pred. No. 47;
  Matches   42;  Conservative   33;  Mismatches   60;  Indels   58;  Gaps    9;

Qy      26 ESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLT 85
           |  |::||  :|     :|   |:     | ||::||| |            ::|   |

Db    1585 EHCLTNFDLAEYRQVLSDLAIQIY-----QQLVRVLENIL----------QPMIVSGMLE 1629

Qy      86 QRIAQDVLRLSSTEPCGLRGCVMHVNLE----IENVCKKLD---RIVCDSSVVPTFEL-- 136
             | |   |  :| |||     :  |    :::: ::|:    ::|   : |  ||
Db    1630 HETIQGV---SGVKPTGLRKRTSSIADEGTYTLDSILRQLNSFHSVMCQHGMDP--ELIK 1684

Qy     137 ----------------TLVFKQENCSW----------TSFRDFFFSRGRFSSGFRRT--- 167
                            |: ::: |||          :   ::   :   :|| : |
Db    1685 QVVKQMFYIVGAITLNNLLLRKDMCSWSKGMQIRYNVSQLEEWLRDKNLMNSGAKETLEP 1744

Qy     168 LILSSGFRLVKKK 180
           || ::    ||||
Db    1745 LIQAAQLLQVKKK 1757


RESULT 13
MY5A_HUMAN
ID   MY5A_HUMAN     STANDARD;      PRT;  1855 AA.
AC   Q9Y4I1; Q9UE30; Q9UE31; Q07902; Q16249; O60653;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)
DE   (Myosin heavy chain 12) (Myoxin).
GN   MYO5A OR MYH12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Skin;
RA   Meurers B.H., Zimmermann R., Vosberg H.P.;
RT   "The complete cDNA for human myosin heavy chain 12, a class V
RT   myosin.";
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.
RX   MEDLINE=97351514; PubMed=9207796;
RA   Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA   Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RT   "Griscelli disease maps to chromosome 15q21 and is associated with
RT   mutations in the myosin-Va gene.";
RL   Nat. Genet. 16:289-292(1997).
RN   [3]
RP   ERRATUM.
RA   Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA   Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RL   Nat. Genet. 23:373-373(1999).
RN   [4]
RP   SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=94245227; PubMed=8188282;
RA   Engle L.J., Kennett R.H.;
RT   "Cloning, analysis, and chromosomal localization of myoxin (MYH12),
RT   the human homologue to the mouse dilute gene.";
RL   Genomics 19:407-416(1994).
RN   [5]
RP   SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=95136715; PubMed=7835087;
RA   Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,
RA   Jenkins N.A.;
RT   "Cloning and regional assignment of the human myosin heavy chain 12
RT   (MYH12) gene to chromosome band 15q21.";
RL   Cytogenet. Cell Genet. 69:53-58(1995).
RN   [6]
RP   SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
RA   Edgar A.J., Bennett J.P.;
RT   "Inhibition of dendrite formation in melanocytes transiently
RT   transfected with antisense DNA to myosin V.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
```

```
RP   FUNCTION.
RX   MEDLINE=99376094; PubMed=10448864;
RA   Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,
RA   Cheney R.E.;
RT   "Myosin-V is a processive actin-based motor.";
RL   Nature 400:590-593(1999).
CC   -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC       STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC       MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC       BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC       FORMATION.
CC   -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC       CALMODULIN OR MYOSIN LIGHT CHAINS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DISEASE: DEFECTS IN MYO5A ARE A CAUSE OF GRISCELLI SYNDROME (GS).
CC       GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN
CC       PIGMENTARY DILUTION OF THE SKIN AND HAIR, THE PRESENCE OF LARGE
CC       CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF
CC       MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN
CC       UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,
CC       KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE
CC       OF BONE MARROW TRANSPLANTATION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U90942; AAD00702.1; -.
DR   EMBL; Y07759; CAA69035.1; -.
DR   EMBL; Y07759; CAA69036.1; -.
DR   EMBL; Z22957; CAA80533.1; -.
DR   EMBL; S74799; AAB33211.1; -.
DR   EMBL; AF055459; AAC14188.1; -.
DR   HSSP; P08799; 1MND.
DR   MIM; 160777; -.
DR   MIM; 214450; -.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00015; IQ; 6.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 6.
KW   Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.
FT   DOMAIN        1    765       MYOSIN HEAD-LIKE.
FT   DOMAIN      766    788       IQ 1.
FT   DOMAIN      789    818       IQ 2.
FT   DOMAIN      814    836       IQ 3.
FT   DOMAIN      837    861       IQ 4.
FT   DOMAIN      862    883       IQ 5.
FT   DOMAIN      885    914       IQ 6.
FT   DOMAIN      914   1237       COILED COIL (POTENTIAL).
FT   DOMAIN     1338   1445       COILED COIL (POTENTIAL).
FT   DOMAIN     1687   1792       DILUTE.
FT   NP_BIND     163    170       ATP (POTENTIAL).
FT   DOMAIN      643    665       ACTIN-BINDING (POTENTIAL).
FT   MOD_RES    1760   1760       PHOSPHORYLATION (POTENTIAL).
FT   VARSPLIC   1321   1347       MISSING (IN ISOFORM 2).
FT   VARSPLIC   1413   1413       L -> LYFEELYADDPKKYQSYRISLYKRMI (IN
FT                                ISOFORM 3).
FT   VARIANT    1246   1246       R -> C.
FT                                /FTId=VAR_010645.
FT   CONFLICT    668    668       F -> L (IN REF. 1).
FT   CONFLICT    833    833       MISSING (IN REF. 4).
FT   CONFLICT    863    863       E -> G (IN REF. 1).
FT   CONFLICT    922    922       H -> R (IN REF. 1).
FT   CONFLICT   1061   1061       V -> L (IN REF. 5).
FT   CONFLICT   1089   1089       E -> Q (IN REF. 4).
FT   CONFLICT   1177   1177       D -> E (IN REF. 5).
FT   CONFLICT   1465   1477       NIPRKEKDFQGML -> SVLCACCVSVTVR (IN REF.
FT                                4).
FT   CONFLICT   1471   1471       K -> N (IN REF. 5).
FT   CONFLICT   1484   1484       E -> D (IN REF. 5).
SQ   SEQUENCE   1855 AA;  215419 MW;  1C55AD57285FA9EC CRC64;

  Query Match             7.7%;  Score 77;  DB 1;  Length 1855;
  Best Local Similarity   21.8%;  Pred. No. 47;
  Matches   42;  Conservative   33;  Mismatches   60;  Indels   58;  Gaps    9;

Qy      26 ESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLT 85
           |  |::||  :|     :|   |:     | ||::||| |           ::|   |
Db    1587 EHCLTNFDLAEYRQVLSDLAIQIY-----QQLVRVLENIL----------QPMIVSGMLE 1631

Qy      86 QRIAQDVLRLSSTEPCGLRGCVMHVNLE----IENVCKKLD---RIVCDSSVVPTFEL-- 136
             | |   |  :| |||     :  |    :::: ::|:    ::|   :  |  ||
Db    1632 HETIQGV---SGVKPTGLRKRTSSIADEGTYTLDSILRQLNSFHSVMCQHGMDP--ELIK 1686

Qy     137 ----------------TLVFKQENCSW----------TSFRDFFFSRGRFSSGFRRT--- 167
                           |: ::: |||          :   ::   :   :|| : |
Db    1687 QVVKQMFYIIGAITLNNLLLRKDMCSWSKGMQIRYNVSQLEEWLRDKNLMNSGAKETLEP 1746

Qy     168 LILSSGFRLVKKK 180
           || ::    ||||
Db    1747 LIQAAQLLQVKKK 1759


RESULT 14
RPOB_MYCGA
ID   RPOB_MYCGA     STANDARD;      PRT;  1390 AA.
AC   P47715;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
DE   beta chain) (RNA polymerase beta subunit).
GN   RPOB.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A5969VAR.B;
RA   Skamrov A.V., Rozovskaya T.A., Goldman M.A., Feoktistova E.S.,
RA   Beabealashvilli R.S.;
RL   Submitted (XXX-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC       ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC       BETA' CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38402; AAB40951.1; -.
DR   InterPro; IPR001572; RNA_pol_B.
DR   Pfam; PF00562; RNA_pol_B; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase.
SQ   SEQUENCE   1390 AA;  155997 MW;  1C02A893FF073542 CRC64;

  Query Match             7.6%;  Score 76;  DB 1;  Length 1390;
  Best Local Similarity   25.4%;  Pred. No. 41;
  Matches   50;  Conservative   36;  Mismatches   91;  Indels   20;  Gaps   11;

Qy      14 SISELLDC-GYHPESLLSDFDYWDYVV----PEPNLNEVIFEESTCQNLVKMLE-NCLSK 67
           ||: ||   |     :   |:  ||::     |    : |  |:    |:: :| : :||
Db     218 SITTLLKAFGLSEIKIKEIFNNNDYILRSLESEFYNEKQILNEADIAQLIRDIESDRISK 277

Qy      68 SKQTKLGCS-KVLVPE--KLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIEN-VCKKLDR 123
             |   :    | || :  || |   |::|  |:  |  :      |: : :   :|:|  :
Db     278 VKSLPIDQKWKNLVLDWYKLNQE-KQELLNSSNPNPTKIESLNTHIGVVLRKLICEKAAK 336

Qy     124 IVCDSSVVPTFELTLVFKQENCSWTS-FRDFFFSRGRF---SSG---FRRTLILSSGF-- 174
           |     : |  |  | ::|  |: |     || : |:   |:|   | | | :|
Db     337 HVIQELSISTRSLDNVAQKEEISYQSILLQHFFQKKRYDLSSAGRHKFVRKLRVSERLYQ 396

Qy     175 RLVKKKLYSLIGTTVIE 191
           | : : :  | |  ||:
Db     397 RTIAQDIKDLDGNVVIK 413


RESULT 15
GP37_HUMAN
ID   GP37_HUMAN     STANDARD;      PRT;   613 AA.
AC   O15354; O14768; O00348;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable G protein-coupled receptor GPR37 precursor (Endothelin B
DE   receptor-like protein-1) (ETBR-LP-1).
GN   GPR37.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97480717; PubMed=9339362;
RA   Marazziti D., Golini E., Gallo A., Lombardi M.S., Matteoni R.,
RA   Tocchini-Valentini G.P.;
RT   "Cloning of GPR37, a gene located on chromosome 7 encoding a putative
RT   G-protein-coupled peptide receptor, from a human frontal brain EST
RT   library.";
RL   Genomics 45:68-77(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Donohue P.J., Shapira H., Mantey S.A., Hampton L.L., Jensen R.T.,
RA   Battey J.F.;
RT   "A human gene encodes a putative G protein-coupled receptor highly
RT   expressed in the central nervous system.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97289672; PubMed=9144577;
RA   Zeng Z., Su K., Kyaw H., Li Y.;
RT   "A novel endothelin receptor type-B-like gene enriched in the brain.";
RL   Biochem. Biophys. Res. Commun. 233:559-567(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Jones K., Tin-Wollam A., Keppler D.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ORPHAN RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12476; CAA73080.1; -.
DR   EMBL; Y12477; CAA73080.1; JOINED.
DR   EMBL; AF017262; AAB70008.1; -.
DR   EMBL; U87460; AAC51281.1; -.
DR   EMBL; AC004925; AAD08853.1; -.
DR   MIM; 602583; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    613       PROBABLE G PROTEIN-COUPLED RECEPTOR
FT                                GPR37.
FT   DOMAIN       27    265       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    266    286       1 (POTENTIAL).
FT   DOMAIN      287    299       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    300    320       2 (POTENTIAL).
FT   DOMAIN      321    335       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    336    356       3 (POTENTIAL).
FT   DOMAIN      357    379       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    380    400       4 (POTENTIAL).
FT   DOMAIN      401    443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    444    464       5 (POTENTIAL).
FT   DOMAIN      465    493       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    494    514       6 (POTENTIAL).
FT   DOMAIN      515    531       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    532    552       7 (POTENTIAL).
FT   DOMAIN      553    613       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      563    568       POLY-CYS.
FT   DISULFID    334    419       BY SIMILARITY.
FT   CARBOHYD     36     36       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     93     93       G -> D (IN REF. 3).
FT   CONFLICT    106    106       A -> T (IN REF. 3).
FT   CONFLICT    118    118       G -> V (IN REF. 3).
FT   CONFLICT    160    160       G -> V (IN REF. 3).
FT   CONFLICT    182    182       W -> C (IN REF. 3).
FT   CONFLICT    231    231       E -> D (IN REF. 3).
FT   CONFLICT    284    284       C -> S (IN REF. 2).
FT   CONFLICT    304    304       A -> V (IN REF. 3).
FT   CONFLICT    329    329       L -> V (IN REF. 3).
FT   CONFLICT    503    504       FC -> LG (IN REF. 3).
FT   CONFLICT    598    598       T -> A (IN REF. 3).
SQ   SEQUENCE   613 AA;  67113 MW;  5A1AB269ED63E765 CRC64;

  Query Match             7.5%;  Score 75.5;  DB 1;  Length 613;
  Best Local Similarity   19.2%;  Pred. No. 17;
  Matches   42;  Conservative   37;  Mismatches   75;  Indels   65;  Gaps    7;

Qy      27 SLLSDFDYWDYVVPEPNLNEVIFEEST------------------------------ 53
           |||::  :||:::    |  ||| | |
Db     298 SLLANLAFWDFLIIFFCLPLVIFHELTKKWLLEDFSCKIVPYIEVASLGVTTFTLCALCI 357

Qy      54 --------CQNLVKMLENCLSKSKQTK---LGCSKVLVPEKLTQRIAQDVLRLSSTEPCG 102
                   |   :|:||| | : :     :|   : :|| :'::::: |  |   |
Db     358 DRFRAATNVQMYYEMIENCSSTTAKLAVIWVGALLLALPEVVLRQLSKEDLGFSGRAPA- 416

Qy     103 LRGCVMHVNLEIENVCKKLDRIVCDSS----------VVPT-FELTLVFKQENCSWTSFR 151
             |:: :: :: :    |  :  ||:           :|| | :|       ||  : |
Db     417 -ERCIIKISPDLPDTIYVL-ALTYDSARLWWYFGCYFCLPTLFTIT-------CSLVTAR 467
```

```
Qy     152 DFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIGTTVI 190
           :    : | :| : | |        |  | |  :|
Db     468 KIRKAEKACTRGNKRQIQLESQMNCTVVALTILYGFCII 506
```

Search completed: October 11, 2002, 14:12:04
Job time : 38 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 11, 2002, 14:02:42 ; Search time 49 Seconds
(without alignments)
378.474 Million cell updates/sec

Title: US-09-881-636-2
Perfect score: 1002
Sequence: 1 MVATGSLSSKNPASISELLD.....FRLVKKKLYSLIGTTVIEGS 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90.5	9.0	1056	2	S56039	RIC1 protein - yea
2	85.5	8.5	711	2	T30107	hypothetical prote
3	85.5	8.5	1414	2	T33236	hypothetical prote
4	82	8.2	673	2	F89101	protein F25E5.7 [i
5	82	8.2	1002	2	G97217	conserved membrane
6	81.5	8.1	228	2	B90292	hypothetical prote
7	80	8.0	927	2	T00357	hypothetical prote
8	79.5	7.9	523	2	D85538	hypothetical prote
9	79.5	7.9	523	2	H90687	hypothetical prote
10	79.5	7.9	606	2	JW0054	amiloride-sensitiv
11	79	7.9	549	2	S72566	cell death protein
12	79	7.9	1808	2	AB1847	serine/threonine k
13	78.5	7.8	463	2	C70363	glycogen synthase
14	78.5	7.8	964	1	T04325	probable ATP-depen
15	77.5	7.7	817	2	D85049	probable transposo
16	77.5	7.7	1153	2	F84468	hypothetical prote
17	77.5	7.7	1584	2	F96573	protein F12M16.25
18	77	7.7	394	2	AG1213	TN916 ORF20 homolo
19	77	7.7	1417	2	A57570	Bloom's syndrome r
20	77	7.7	1828	2	B59254	mysoin heavy chain
21	77	7.7	1830	1	S19188	myosin-V - chicken
22	77	7.7	1853	1	A46761	myosin heavy chain
23	77	7.7	1855	2	A59254	mysoin heavy chain
24	76.5	7.6	613	2	JC5501	endothelin recepto
25	76.5	7.6	1309	2	H96650	protein T3P18.3 [i
26	76	7.6	1083	2	S59780	hypothetical prote
27	76	7.6	1871	2	D96698	probable DNA polym
28	76	7.6	1894	2	T02155	DNA-directed DNA p
29	75.5	7.5	963	2	T26022	hypothetical prote
30	75.5	7.5	1603	2	D89407	protein R10E8.6 [i
31	75.5	7.5	1696	2	T24146	hypothetical prote
32	75	7.5	509	2	F71526	hypothetical prote
33	75	7.5	1819	2	A71928	cag island protein
34	74.5	7.4	979	2	T41006	ubiquitin carboxyl
35	74.5	7.4	1171	2	T05039	hypothetical prote
36	74.5	7.4	1889	2	T50069	dynein heavy chain
37	74.5	7.4	4196	2	T43274	dynein heavy chain
38	74	7.4	452	2	T09596	cyclin cyc3 - alfa
39	74	7.4	585	2	E85040	hypothetical prote
40	74	7.4	1405	2	T04426	hypothetical prote
41	73.5	7.3	366	2	S25277	flagella-associate
42	73.5	7.3	453	2	D85683	hypothetical prote
43	73.5	7.3	453	2	H90824	hypothetical prote
44	73.5	7.3	949	2	T38543	hypothetical prote
45	73.5	7.3	1441	2	T13889	helicase II homolo

ALIGNMENTS

RESULT 1
S56039
RIC1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein L1917; protein YLR039c
C;Species: Saccharomyces cerevisiae
C;Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 20-Jun-2000
C;Accession: S56039; S64866
R;Mizuta, K.
submitted to the EMBL Data Library, December 1994
A;Reference number: S56039
A;Accession: S56039
A;Molecule type: DNA
A;Residues: 1-1056 <MIZ>
A;Cross-references: EMBL:D43895; NID:g619062; PID:g633123
R;Koetter, P.; Rose, M.; Entian, K.D.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64863
A;Accession: S64866
A;Molecule type: DNA
A;Residues: 1-1056 <KOE>
A;Cross-references: EMBL:Z73211; NID:g1360365; PID:g1360366; MIPS:YLR039c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:RIC1
A;Cross-references: SGD:S0004029; MIPS:YLR039c
A;Map position: 12R
C;Superfamily: Saccharomyces cerevisiae RIC1 protein

Query Match 9.0%; Score 90.5; DB 2; Length 1056;
Best Local Similarity 25.0%; Pred. No. 4;
Matches 42; Conservative 23; Mismatches 52; Indels 51; Gaps 8;

```
Qy     24 HPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEK 83
          :|:  |||:  |:| |||   | ::|          :  | ||: |  |:        ||
Db    613 NPDVELSDYVMWEYNVPE---NTIVF---------SLHVNTLSRYKLLKMKSKNHNASEK 660

Qy     84 L---------------TQRIAQDVLRLSSTEPCGLRGC-----VMHVNLEIENVCKKLDR 123
                           || |  ||  :|:  ||||         : :|: |: :  |::
Db    661 QPDALLKTAEIILVTDTQTIVFDV--ISTVHPCGLNIIKKFYQYLKINIPIDVLPNKIEW 718

Qy    124 IVCDSSVVPTFELTLVFKQENCSWTSFRD-FFFSRGRFSSGFRRTLIL 170
          |:      :  |:|       | |  |  |:  |  :|| |
Db    719 II-------NMKEGLLF---------FADRKFIKLGKVDGGGWQTLTL 750
```

RESULT 2
T30107
hypothetical protein F44C4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T30107
R;Geisel, C.; Bradshaw, H.

```
submitted to the EMBL Data Library, February 1996
A;Description: The sequence of C. elegans cosmid F44C4.
A;Reference number: Z20736
A;Accession: T30107
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-711 <GEI>
A;Cross-references: EMBL:U50313; PIDN:AAA92330.1; CESP:F44C4.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:F44C4.1
A;Introns: 56/3; 76/1; 101/1; 161/1; 196/1; 226/3; 269/2; 318/1; 364/1; 475/3; 520/2; 56

  Query Match             8.5%;  Score 85.5;  DB 2;  Length 711;
  Best Local Similarity   23.6%;  Pred. No. 7.1;
  Matches   46;  Conservative   35;  Mismatches   73;  Indels   41;  Gaps   10;

Qy      16 SELLDCGYHPESLLSDF---DY---WDYVVPEPNLNEVIFEESTCQNLVKM--------- 60
           |: :||| :    : ||   ||   |  |    : :| : | :|   | |:
Db     358 SDEMDCG-NINGTMCDFNGQDYCNSWYQVTNVTDYHERLSEPTTVAPLNKLNEVPLHLFR 416

Qy      61 LENCLSKSKQTKLGCSKVLV----PEKLTQRIAQDVL-RLSSTEPCG-------LRGCVM 108
           |::  :| |:   |   :||    |  ||:| :  |   |  | |           : | :
Db     417 LQSPSAKIKEAMRGSGNMLVFDHKPNPLTRRTSALVSPELPRTNPEAYDEKSPLFKSCKL 476

Qy     109 HVNL---------EIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGR 159
             |         :|  : | :: :    :::     ||: | |||:|      | :  |
Db     477 RFYLCSRTYSKVWQISVISKGINPMESGRTIIYEAGYTLIPK-ENCTW---ERVFVNIPR 532

Qy     160 FSSGFRRTLILSSGF 174
           ::|||  : ::: |
Db     533 QNAGFRIGIFVTNYF 547


RESULT 3
T33236
hypothetical protein T10H9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33236
R;Greco, T.; Bradshaw, H.; O'Brien, D.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid T10H9.
A;Reference number: Z21306
A;Accession: T33236
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1414 <GRE>
A;Cross-references: EMBL:AF067949; PIDN:AAC19236.1; GSPDB:GN00023; CESP:T10H9.2
A;Experimental source: strain Bristol N2; clone T10H9
C;Genetics:
A;Gene: CESP:T10H9.2
A;Map position: 5
A;Introns: 56/3; 76/1; 101/1; 161/1; 196/1; 226/3; 269/2; 318/1; 364/1; 475/3; 520/2; 56

  Query Match             8.5%;  Score 85.5;  DB 2;  Length 1414;
  Best Local Similarity   23.6%;  Pred. No. 16;
  Matches   46;  Conservative   35;  Mismatches   73;  Indels   41;  Gaps   10;

Qy      16 SELLDCGYHPESLLSDF---DY---WDYVVPEPNLNEVIFEESTCQNLVKM--------- 60
           |: :||| :    : ||   ||   |  |    : :| : | :|   | |:
Db     358 SDEMDCG-NINGTMCDFNGQDYCNSWYQVTNVTDYHERLSEPTTVAPLNKLNEVPLHLFR 416

Qy      61 LENCLSKSKQTKLGCSKVLV----PEKLTQRIAQDVL-RLSSTEPCG-------LRGCVM 108
           |::  :| |:   |   :||    |  ||:| :  |   |  | |           : | :
Db     417 LQSPSAKIKEAMRGSGNMLVFDHKPNPLTRRTSALVSPELPRTNPEAYDEKSPLFKSCKL 476

Qy     109 HVNL---------EIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGR 159
             |         :|  : | :: :    :::     ||: | |||:|      | :  |
Db     477 RFYLCSRTYSKVWQISVISKGINPMESGRTIIYEAGYTLIPK-ENCTW---ERVFVNIPR 532

Qy     160 FSSGFRRTLILSSGF 174
           ::|||  : ::: |
Db     533 QNAGFRIGIFVTNYF 547


RESULT 4
F89101
protein F25E5.7 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: F89101
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: F89101
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-673 <STO>
A;Cross-references: GB:chr_V; PIDN:AAC27338.1; PID:g3335245; GSPDB:GN00023; CESP:F25E
A;Note: weak similarity to coagulation factor IX
C;Genetics:
A;Gene: F25E5.7
A;Map position: 5

  Query Match             8.2%;  Score 82;  DB 2;  Length 673;
  Best Local Similarity   26.9%;  Pred. No. 14;
  Matches   29;  Conservative   21;  Mismatches   36;  Indels   22;  Gaps    5;

Qy       2 VATGSLSSKNPASISEL--------LDCGYHPESLLSDFDYWDYVVPEPNLN--EVIFEE 51
           :| | |:: :|  : ::         :|  |||:  | :  | :      |  ||:
Db     185 LARGRLNTADPYILEDITPIDFKLYMDLTYHPKQYFSAYHAKDILQIAKRFNNLEVV--- 241

Qy      52 STCQNLV--KMLENCLSKSKQTKLGCS-------KVLVPEKLTQRIAQ 90
           :||||::   : || ||   :|||  |       ::  |  :||  |:
Db     242 TTCQNVILEDLNENKLSTKNKTKLAESYDLDTVRELFNPNTMTQPSAR 289


RESULT 5
G97217
conserved membrane protein CAC2581 (6-pyruvoyl-tetrahydropterin synthase related doma
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: G97217
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97217
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1002 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80530.1; PID:g15025604; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2581

  Query Match             8.2%;  Score 82;  DB 2;  Length 1002;
  Best Local Similarity   19.0%;  Pred. No. 23;
  Matches   48;  Conservative   40;  Mismatches   83;  Indels   82;  Gaps   10;

Qy       5 GSLSSKNPASISELLDCGYHP----------------------ESLLSDFDYW----D 36
           |: :| |   ::  |: ||:                        || ||
Db     625 GAATSSNIVLLNTALENGYYKYMFDRSLEMGCDTVVLRKSVIKDNRKSFSDIDYGAKLSG 684

Qy      37 YVVPEPNLNEVIFEESTCQN---LVKMLENCLSKSKQT--------KLGCSKVLVPEKLT 85
           | :   | :  || :|| :    : |    |: || :          |:| |  |    |
Db     685 YKLCFENADSYIFHKSTPKAFGVITKYEGLCIGKSAKEIQLLYPSFKIGSSNNLEDYSLN 744

Qy      86 QRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLD----RIVCDSSVVPTFELT---- 137
             :   :: ||        |   :   : | : |||     |:: | : :||  :|
Db     745 DLMKYKIIYLS--------GFSYNNRSKAEGIVKKLSKKGVRVIVDMNRIPTDNITDRMI 796
```

```
Qy    138 --------LVFKQENCSWTSFRDFFFSRGRFSSGFR--------RTLILSS----GFRL 176
                  :||  :        | ||:: |: :  :          |: |::     ||
Db    797 FLGVISEPIVFDNK------MPDLFFNKERYVADVKFKEEYKKWNTVYLTNTQNVKGFSW 850

Qy    177 VKKKLYSLIGTTV 189
          :  |  : : : :
Db    851 INDKKIAFLASGI 863


RESULT 6
B90292
hypothetical protein SSO1357 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: B90292
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-228 <KUR>
A;Cross-references: GB:AE006641; NID:g13814569; PIDN:AAK41593.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO1357

  Query Match             8.1%;  Score 81.5;  DB 2;  Length 228;
  Best Local Similarity   22.8%;  Pred. No. 4.2;
  Matches   31;  Conservative   24;  Mismatches   48;  Indels   33;  Gaps    5;

Qy     53 TCQNLVKMLENCLSKSKQTKLGCSKVLVPE-KLTQRIAQDVLRLSSTEPCGLRGCVMHVN 111
          ||:  :||:                |::|     :|| : : ::||| |          |
Db     17 TCKESLKMI--------------SVIIPAFNEERRIGKTLEKISSTLP----------N 51

Qy    112 LEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILS 171
           |:  |    |       ||  | : |:  :|       :     ||   | |:| |:|
Db     52 AEVVVVFDGHDN---TPEVVKKFPVKLIISKERLG----KGMALKRGITESNFQRVLLLD 104

Qy    172 SGFRLVKKKLYSLIGT 187
          : | : :::|  :: |
Db    105 ADFPITEEELNKILST 120


RESULT 7
T00357
hypothetical protein KIAA0685 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00357
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A;Reference number: Z14142; MUID:98403880
A;Accession: T00357
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-927 <ISH>
A;Cross-references: EMBL:AB014585; NID:g3327183; PIDN:EAA31660.1; PID:g3327184
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0685

  Query Match             8.0%;  Score 80;  DB 2;  Length 927;
  Best Local Similarity   32.0%;  Pred. No. 32;
  Matches   41;  Conservative   18;  Mismatches   53;  Indels   16;  Gaps    7;

Qy     10 KNPASISELLDCGY--------HPESLLSDFDYWDYVVPEPNLNEVI--FEESTCQNLV- 58
          | | :  ||| |            |||||    :|::  || || ::  |  |  ||:
Db     76 KYPNTACELLTCDVPQISDRLGGDESLLSLL--YDFLDHEPPLNPLLASFFSKTIGNLIA 133

Qy     59 KMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRL-SSTEPCGLRGCVMHVNLEIENV 117
          :  |  ::  |:     | ||     |  :   :||| |  || |||  |:|  |  | |
Db    134 RKTEQVITFLKKKDKFISLVL-KHIGTSALMDLLLRLVSCVEPAGLRQDVLHW-LNEEKV 191

Qy    118 CKKLDRIV 125
           ::|  ::
Db    192 IQRLVELI 199


RESULT 8
D85538
hypothetical protein Z0521 [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85538
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-523 <STO>
A;Cross-references: GB:AE005174; NID:g12513274; PIDN:AAG54768.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0521

  Query Match             7.9%;  Score 79.5;  DB 2;  Length 523;
  Best Local Similarity   24.5%;  Pred. No. 18;
  Matches   40;  Conservative   24;  Mismatches   62;  Indels   37;  Gaps    7;

Qy     29 LSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS-KSKQTKLGC--SKVL--VPEK 83
          ||  :|   : |:|| |  |   |   : :|:  : || | ||| | |  :|:  :||
Db     63 LSVLNYNSALTPDPNFN--IDATSDLDSYLKLDFDRLSPKQKQTTLCCFWNKIASSLPEP 120

Qy     84 LTQRIAQDVLRLSSTEPCGLRGCVMHVN-------LEIE----------------NVCK 119
              |  :::     |   :|| :  ||        | ||                  |:
Db    121 YNSTIKHNIIFYKDGENLMIRGTISIVNEVVKTYSLPIEKDDNGYYDFSGLYLAHSNISG 180

Qy    120 KLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSS 162
          |      | :  |  :  :     |||  :|   :|   :|::
Db    181 K------DPNKDPDIDFGIDMGNCNCSNVNFEHTYFYGVKFTN 217


RESULT 9
H90687
hypothetical protein ECs0472 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H90687
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90687
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-523 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33895.1; PID:g13359929; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0472

  Query Match             7.9%;  Score 79.5;  DB 2;  Length 523;
  Best Local Similarity   24.5%;  Pred. No. 18;
  Matches   40;  Conservative   24;  Mismatches   62;  Indels   37;  Gaps    7;

Qy     29 LSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLS-KSKQTKLGC--SKVL--VPEK 83
          ||  :|   : |:|| |  |   |   : :|:  : || | ||| | |  :|:  :||
```

```
Db      63 LSVLNYNSALTPDPNFN--IDATSDLDSYLKLDFDRLSPKQKQTTLCCFWNKIASSLPEP 120

Qy      84 LTQRIAQDVLRLSSTEPCGLRGCVMHVN-------LEIE-----------------NVCK 119
           |  :::     |   :|| :  ||       | ||                   |:
Db     121 YNSTIKHNIIFYKDGENLMIRGTISIVNEVVKTYSLPIEKDDNGYYDFSGLYLAHSNISG 180

Qy     120 KLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSS 162
           |      | :  |  :  :     |||  :|   :|   :|::
Db     181 K------DPNKDPDIDFGIDMGNCNCSNVNFEHTYFYGVKFTN 217


RESULT 10
JW0054
amiloride-sensitive sodium channel protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Nov-2000
C;Accession: JW0054
R;Darboux, I.; Lingueglia, E.; Pauron, D.; Barbry, P.; Lazdunski, M.
Biochem. Biophys. Res. Commun. 246, 210-216, 1998
A;Title: A new member of the amiloride-sensitive sodium channel family in Drosophila mel
A;Reference number: JW0054; MUID:98262943
A;Accession: JW0054
A;Molecule type: mRNA
A;Residues: 1-606 <DAR>
A;Cross-references: GB:Y16225; NID:g2791943; PIDN:CAA76124.1; PID:g2791944
C;Genetics:
A;Cross-references: FlyBase:FBgn0020258
A;Introns: 91/2 179/2 349/3 396/2 439/3 520/2
F;92-120/Domain: transmembrane #status predicted <TM1>
F;558-575/Domain: transmembrane #status predicted <TM2>

  Query Match             7.9%;  Score 79.5;  DB 2;  Length 606;
  Best Local Similarity   19.3%;  Pred. No. 21;
  Matches   35;  Conservative   32;  Mismatches   47;  Indels   67;  Gaps    9;

Qy      37 YVVPEPNLNEVI------------------FEESTCQNLVKMLENCLSKSKQTKLGCSK 77
           |||   ||:|:                  : :| ||       ||:    :| ||:|
Db     382 YVVSTDNLHEITPEKRQCLFDDERSLRFFRSYSQSNCQT------ECLANYTVSKCGCAK 435

Qy      78 VLVPEKLTQRIAQDVLRLSSTEPCGLR--GCV------MHVNLEIENVCKKLDR---IVC 126
           :|: |              |  |||:   |       ::  :: : : | :|    | |
Db     436 FWMPKPL------------GTPVCGLKDINCYTSAQDELYTLMQNQTMAKSIDESVDITC 483

Qy     127 D-----SSVVPTFELTLVFKQENCSWTSFRD------------FFFSRGRFSSGFRRTL 168
           :     :|:   ||::        :  :||:              :|   :|::  :||:
Db     484 NCMPACTSLEYNFEISRAKYDVAKTIRAFREEYEHTDAIGSRLSVYFKEHQFTA-IKRTI 542

Qy     169 I 169
           :
Db     543 L 543


RESULT 11
S72566
cell death protein ced-4 - Caenorhabditis elegans
N;Alternate names: hypothetical protein C35D10.9
C;Species: Caenorhabditis elegans
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 31-Oct-1997
C;Accession: S72566; A49152
R;Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton,
ulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnston, L.; Jon
, B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.
submitted to the EMBL Data Library, February 1995
A;Authors: Shownkeen, R.; Smaldon, N.; Smith, A.; Sonnhammer, E.; Staden, R.; Sulston, J
proat, J.; Wohldman, P.
A;Description: The C. elegans genome project: Contiguous nucleotide sequence of over two
A;Reference number: S72566
A;Accession: S72566
A;Molecule type: DNA
A;Residues: 1-549 <WIL>
A;Cross-references: EMBL:U21324; NID:g687879; PID:g687887
A;Experimental source: strain Bristol N2
R;Yuan, J.; Horvitz, H.R.
Development 116, 309-320, 1992
A;Title: The Caenorhabditis elegans cell death gene ced-4 encodes a novel protein and
A;Reference number: A49152; MUID:93161939
A;Accession: A49152
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-211,213-549 <YUA>
A;Note: sequence extracted from NCBI backbone (NCBIN:124821, NCBIP:124822)
C;Genetics:
A;Gene: ced-4
A;Map position: 3
A;Introns: 151/1; 182/2; 212/2; 234/3; 347/2; 439/2; 474/3

  Query Match             7.9%;  Score 79;  DB 2;  Length 549;
  Best Local Similarity   24.7%;  Pred. No. 21;
  Matches   45;  Conservative   23;  Mismatches   46;  Indels   68;  Gaps   11;

Qy      10 KNPASISELLD-CGYHPESLLSDF--DYWDYVVPEPNLNE--VIFEESTCQNLVKMLENC 64
           :  : :  |:|   |: :| |:||  || |: : ||:|    ||  : : | |
Db      63 RQASELGPLIDFFNYNNQSHLADFLEDYIDFAINEPDLLRPVVIAPQFSRQML------- 115

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
              ::  ||     ||:::|               | :|         :: | ||||
Db     116 ---DRKLLLG----NVPKQMT--------------CYIR------EYHVDRVIKKLD-- 145

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSL 184
                           | |   ||  | |  ||  ||  :::| |        |   |
Db     146 ------------------EMCDLDSF--FLFLHGRAGSG--KSVIASQAL----SKSDQL 179

Qy     185 IG 186
           ||
Db     180 IG 181


RESULT 12
AB1847
serine/threonine kinase with two-component sensor domain all0323 [imported] - Anabaen
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AB1847
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB1847
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1808 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB72281.1; PID:g17129668; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all0323

  Query Match             7.9%;  Score 79;  DB 2;  Length 1808;
  Best Local Similarity   20.2%;  Pred. No. 89;
  Matches   41;  Conservative   36;  Mismatches   68;  Indels   58;  Gaps   10;

Qy      11 NPASISELLDCGYHPESLLSDF--DYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKS 68
           ||  |::|||   |    |: ||    | :                    : |::   :|
Db     573 NPFFINQLLKALYQESLLIFDFTTQKWQW-----------------EIEKIITVGVSDK 614

Qy      69 KQTKLGCSKV-LVPEKLTQRIAQDVLRLSSTEPCGLRGCV-----MHV-----NLEIENV 117
             :|  |:: |:||| ||:| |          |  |:     :||       : :
Db     615 SVVELVSSRIQLLPEK-TQKILQ-----------LAACIGDKFNLHVLSLVNQKSVFDT 661

Qy     118 CKKLDRIVCDSSVVP---TFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGF 174
             :||  :  | ::|   |: : |:|  ::       |  |   :    |    :  : :
Db     662 ANELDDALQSSLILPLNETYRIPLLFHSQD-------DISFDSKQILYKFLHDRVQQAAY 714

Qy     175 RLV----KKKLYSLIGTTVIEGS 193
           |:    ||  :  ||  ::: :
```

```
Db    715 SLIPQNQKKDTHIKIGRLLLQNT 737


RESULT 13
C70363
glycogen synthase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C;Accession: C70363
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70363
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-463 <AQF>
A;Cross-references: GB:AE000704; NID:g2983301; PIDN:AAC06894.1; PID:g2983303; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: glgA
C;Superfamily: starch synthase

  Query Match             7.8%;  Score 78.5;  DB 2;  Length 463;
  Best Local Similarity   19.6%;  Pred. No. 19;
  Matches   37;  Conservative   37;  Mismatches   78;  Indels   37;  Gaps    5;

Qy      1 MVATGSLSSKNPASISEL--LDCGYHPESLLSDFDY--------WDYVVPEPNLNEVIFE 50
          :| :  ::: :|    |:   : ||  | :|  : |         || |  |  :: |::
Db    198 IVFSDLITTVSPTYAKEIQTQEYGYGLEGVLKKYSYKLRGILNGIDYEVWNPEKDKYIYQ 257

Qy     51 ESTCQNLVKMLENCLSKSKQTKLGCSKVLVP--EKLTQRIAQDVLRLSSTEPCGLRGCVM 108
            : :|  |  :|    ||:  :   | |:    : | :   ::
Db    258 NYSLRNYSKKFKNKEFLSKELGIEAEKPLISFINRFTHQKGVEL--------------- 301

Qy    109 HVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTL 168
              | |  ::: ::  :   : | |    |   :  : :|: |      |: || | |
Db    302 -----ILNCAEEMSKLNANFVFLGTGEYENAFLDVSKIYKNFKVF----AEFNEGFARKL 352

Qy    169 ILSSGFRLV 177
            || | |:
Db    353 YASSDFILM 361


RESULT 14
T04325
probable ATP-dependent proteinase LON2 (EC 3.4.21.-), mitochondrial - maize
C;Species: Zea mays (maize)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-Jan-2001
C;Accession: T04325
R;Barakat, S.; Pearce, D.A.; Sherman, F.; Rapp, W.D.
Plant Mol. Biol. 37, 141-154, 1998
A;Title: Maize contains a Lon protease gene that can partially complement a yeast pim1-d
A;Reference number: Z15282; MUID:98281582
A;Accession: T04325
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-964 <BAR>
A;Cross-references: EMBL:U85495; NID:g1816587; PIDN:AAC50021.1; PID:g1816588
A;Experimental source: strain B73
C;Genetics:
A;Gene: LON2
C;Function:
A;Description: serine proteinase
C;Superfamily: ATP-dependent serine proteinase La
C;Keywords: ATP; DNA binding; hydrolase; mitochondrial matrix; mitochondrion; molecular
F;455-462/Region: nucleotide-binding motif A (P-loop)
F;518-523/Region: nucleotide-binding motif B
F;461/Binding site: ATP (Lys) #status predicted
F;863/Active site: Ser #status predicted

  Query Match             7.8%;  Score 78.5;  DB 1;  Length 964;
  Best Local Similarity   27.6%;  Pred. No. 46;
  Matches   34;  Conservative   28;  Mismatches   42;  Indels   19;  Gaps    8;

Qy      5 GSLSSKNPAS-ISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTC-QNLVKMLE 62
          |   | :||| : ||||    ||  ::  |:  |:    :|::|:|    |  |:::|:
Db    528 GKGHSGDPASALLELLD----PEQNVNFLDH--YLDVPIDLSKVLF---VCTANVIEMIP 578

Qy     63 NCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVN-----LEIENV 117
          | |   :   :  :  :  ||:   ||:| |  :: : ||::   : |        |||
Db    579 NPL-LDRMEIIAIAGYITDEKM--HIARDYLEKNTRQACGIKPEQVEVTDTALLALIENY 635

Qy    118 CKK 120
          |::
Db    636 CRE 638


RESULT 15
D85049
probable transposon protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: D85049
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: D85049
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-817 <STO>
A;Cross-references: GB:NC_001268; NID:g7267145; PIDN:CAB80813.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g03900
A;Map position: 4

  Query Match             7.7%;  Score 77.5;  DB 2;  Length 817;
  Best Local Similarity   20.1%;  Pred. No. 47;
  Matches   36;  Conservative   29;  Mismatches   71;  Indels   43;  Gaps    5;

Qy      7 LSSKNPASISELLDCG------------YHPESLLSDFDYWDYVVPEPNLNEVIFEEST 53
          |:| ||    ::   |              :| ||:| :  ||  :    |:: :  |::
Db    540 LASVNPPKTKDVGGNGHEKKMRGYGKEHNWHKESILWELSYWKDLNLRHNIDVMHTEKNF 599

Qy     54 CQNLVKMLENCLSKSKQTKLG-------CSK-----------VLVPEKLTQRIAQDVLRL 95
            |::  |     |||   :         ||:             | |||:   | : :
Db    600 LDNIMNTLLGVKGKSKDNIMSRLDIEKYCSRPGLHIDSTGKAPFPPYKLTEEAKQSLFQ- 658

Qy     96 SSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFF 154
                || | ::   :  |  |  |    ::|     |: :  : : :  | ||
Db    659 ----------CVKH-DVRFPDGMKSHDCHVFMERLLPFIFAELLDRNVHLALSGIRAFF 706


Search completed: October 11, 2002, 14:14:48
Job time : 52 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 11, 2002, 12:29:47 ; Search time 63 Seconds
(without alignments)
340.274 Million cell updates/sec

Title: US-09-881-636-2
Perfect score: 1002
Sequence: 1 MVATGSLSSKNPASISELLD.....FRLVKKKLYSLIGTTVIEGS 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	272.5	27.2	177	20	AAY59820	Human normal ovari
2	271	27.0	232	20	AAY03635	Hypoxia-regulated
3	271	27.0	232	22	AAE03922	Human gene 25 enco
4	271	27.0	233	21	AAB51661	Human secreted pro
5	261	26.0	229	20	AAY03634	Hypoxia-regulated
6	211.5	21.1	299	22	ABB63344	Drosophila melanog
7	193	19.3	280	22	ABB63379	Drosophila melanog
8	112.5	11.2	74	21	AAB45148	Human secreted pro
9	85	8.5	1066	22	ABB64251	Drosophila melanog
10	82	8.2	934	22	AAB60490	Human cell cycle a
11	80.5	8.0	498	22	AAB92900	Human protein sequ
12	79.5	7.9	606	22	ABB59933	Drosophila melanog
13	79.5	7.9	860	22	AAM79768	Human protein SEQ
14	79	7.9	169	22	AAU20460	Human secreted pro
15	79	7.9	261	15	AAR53277	Ced-4 (Q262X). Ca
16	79	7.9	315	21	AAB43007	Human ORFX ORF2771
17	79	7.9	315	21	AAY84374	A human a neuronal
18	79	7.9	322	22	ABB11760	Human voltage gate
19	79	7.9	409	15	AAR53278	Ced-4 (W401X(sic))
20	79	7.9	549	15	AAR53276	Ced-4 (I258N). Ca
21	79	7.9	549	15	AAR47465	Ced-4. Caenorhabd
22	78	7.8	179	22	AAU31632	Novel human secret
23	78	7.8	224	22	ABG05646	Novel human diagno
24	78	7.8	224	22	ABG05652	Novel human diagno
25	78	7.8	371	22	AAU28240	Novel human secret
26	78	7.8	481	22	ABG15548	Novel human diagno
27	78	7.8	1056	22	ABG09490	Novel human diagno
28	77.5	7.7	613	21	AAY90667	Human mutant G pro
29	77	7.7	515	18	AAW31547	Bloom's syndrome B
30	77	7.7	739	18	AAW31549	Bloom's syndrome B
31	77	7.7	1417	18	AAW31548	Bloom's syndrome B
32	77	7.7	1417	18	AAW31550	Bloom's syndrome B
33	77	7.7	1417	18	AAW31551	Bloom's syndrome B
34	77	7.7	1418	18	AAW15264	Bloom syndrome act
35	76.5	7.6	314	21	AAY84375	Protein encoded by
36	76.5	7.6	613	17	AAR98261	Human endothelin-b
37	76.5	7.6	613	19	AAW37799	Amino acid sequenc
38	76.5	7.6	1262	22	ABB59586	Drosophila melanog
39	76	7.6	1364	22	AAG70912	S cerevisiae apopt
40	75.5	7.5	348	22	ABG05997	Novel human diagno
41	75.5	7.5	613	21	AAY90633	Human G protein-co
42	75	7.5	446	20	AAY36840	Amino acid sequenc
43	75	7.5	748	22	AAU07878	Polypeptide sequen
44	74.5	7.4	323	21	AAY84376	A human voltage-ga
45	74.5	7.4	325	22	ABB11805	Human voltage gate

ALIGNMENTS

RESULT 1
AAY59820
ID AAY59820 standard; Protein; 177 AA.
XX
AC AAY59820;
XX
DT 18-JAN-2000 (first entry)
XX
DE Human normal ovarian tissue derived protein 97.
XX
KW Human; ovary; screening; ovarian cancer; treatment.
XX
OS Homo sapiens.
XX
PN DE19816395-A1.
XX
PD 07-OCT-1999.
XX
PF 03-APR-1998; 98DE-1016395.
XX
PR 03-APR-1998; 98DE-1016395.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-552352/47.
DR N-PSDB; AAZ41261.
XX
PT Nucleic acid sequences potentially useful in diagnosis or therapy of
PT ovarian cancer -
XX
PS Claim 23; Page 255; 274ppp; German.
XX

```
CC   This invention describes novel nucleic acid sequences that are highly
CC   expressed in normal ovary tissue. Artificial chromosomes and cosmid
CC   clones containing the sequences can be used as gene transfer vehicles.
CC   The sequences can be used to produce DNA fragments containing
CC   full-length genes. Host cells transformed with the sequences can be used
CC   to produce polypeptides or polypeptide fragments, which can be used to
CC   screen phage displays for polypeptides that bind to them, or as tools for
CC   identifying agents active against ovarian cancer, or to prepare
CC   medicaments for treating ovarian cancer. The cDNA sequences can be used
CC   to obtain genomic genes, their promoters, enhancers, silencers, exon
CC   structures, intron structures and their splice variants. AAY59724-Y59837
CC   represent protein fragments encoded by the cDNA sequences represented in
CC   AAZ41222-Z41324 which are derived from normal human ovarian tissue.
XX
SQ   Sequence   177 AA;

  Query Match             27.2%;  Score 272.5;  DB 20;  Length 177;
  Best Local Similarity   33.0%;  Pred. No. 9.6e-23;
  Matches   58;  Conservative   44;  Mismatches   63;  Indels   11;  Gaps    3;

Qy       9 SKNPASISELLDCGYHPESLLSDFDYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKS 68
           |  |: :    |  |     | ||:    :: :|       :|  | ||:::|:  |::: 
Db       3 SNTPSCVLSFPDTAYLDGVSLPDFE----LLSDPE------DEHLCANLMQLLQESLAQA 52

Qy      69 KQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDS 128
           :      :::|:| :|  :: :::|||: :|||||| :: | :|    |  : ::  |
Db      53 RLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQLALDP 112

Qy     129 SVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
           |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::|||||| 
Db     113 SLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 168


RESULT 2
AAY03635
ID   AAY03635 standard; Protein; 232 AA.
XX
AC   AAY03635;
XX
DT   04-JUN-1999  (first entry)
XX
DE   Hypoxia-regulated gene RTP779 product.
XX
KW   Hypoxia-regulated gene; therapeutic; diagnostic; hypoxia; ischemia;
KW   apoptosis; angiogenesis; tumorigenic cell; trauma; limb reattachment;
KW   revascularisation.
XX
OS   Rattus sp.
XX
PN   WO9909049-A1.
XX
PD   25-FEB-1999.
XX
PF   21-AUG-1998;   98WO-US17296.
XX
PR   21-AUG-1997;   97US-0056453.
XX
PA   (KOHN/) KOHN K I.
PA   (QUAR-) QUARK BIOTECH INC.
XX
PI   Einat P,  Skaliter R;
XX
DR   WPI; 1999-180965/15.
DR   N-PSDB; AAX29138.
XX
PT   New isolated hypoxia-related genes  - used to develop products for
PT   use in therapy and diagnosis in e.g. hypoxia, ischaemia, apoptosis
PT   and angiogenesis
XX
PS   Claim 14; Page 82-83; 92pp; English.
XX
CC   Sequences AAY03632-36 represent products of hypoxia-regulated genes of
CC   the invention. The genes and their products can be used therapeutically
CC   and diagnostically in hypoxia, ischemia, apoptosis and angiogenesis. The
CC   products and methods can be used for e.g. inducing apoptosis in
CC   tumorigenic cells or angiogenesis in trauma situations where e.g. a limb
CC   must be reattached or in a transplant where revascularisation is needed.
XX
SQ   Sequence   232 AA;

  Query Match             27.0%;  Score 271;  DB 20;  Length 232;
  Best Local Similarity   35.0%;  Pred. No. 2.1e-22;
  Matches   63;  Conservative   45;  Mismatches   66;  Indels    6;  Gaps    4;

Qy       6 SLSSKNPASISELLDCGYHPESLLSDFDYWDYV-VPEPNLNEVIFEESTCQNLVKMLENC 64
           || | :  |: :  : |: ||    |  | | | :|:  |     :|  | ||:::|:
Db      48 SLESSDCESL-DSSNSGFGPE---EDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQES 103

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           |::::      :::|:| :|  :: :::|||: :|||||| :: | :|    |  : ::
Db     104 LAQARLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQL 163

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             | |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::|||||| 
Db     164 ALDPSLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 223


RESULT 3
AAE03922
ID   AAE03922 standard; Protein; 232 AA.
XX
AC   AAE03922;
XX
DT   09-AUG-2001  (first entry)
XX
DE   Human gene 25 encoded secreted protein HUKCS86, SEQ ID NO:85.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal_peptide
FT   Protein         22..232
FT                   /note= "Mature secreted protein"
XX
PN   WO200077022-A1.
XX
PD   21-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15136.
XX
PR   11-JUN-1999;   99US-0138629.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-367020/38.
DR   N-PSDB; AAD08369.
XX
PT   Nucleic acids encoding 50 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating diseases, e.g. Parkinson's
PT   disease, botulism, cancers and Scimitar syndrome -
XX
```

```
PS   Claim 11; Page 538-539; 614pp; English.
XX
CC   AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted
CC   protein genes and AAE03898-AAE03947 represent the proteins they encode.
CC   AAE03948-AAE03996 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   50 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein of the invention.
XX
SQ   Sequence   232 AA;

  Query Match             27.0%;  Score 271;  DB 22;  Length 232;
  Best Local Similarity   35.0%;  Pred. No. 2.1e-22;
  Matches   63;  Conservative   45;  Mismatches   66;  Indels    6;  Gaps    4;

Qy       6 SLSSKNPASISELLDCGYHPESLLSDFDYWDYV-VPEPNLNEVIFEESTCQNLVKMLENC 64
           || | :  |: :  : |: ||    |  | | | :|:  |     :|  | ||:::|:
Db      48 SLESSDCESL-DSSNSGFGPE---EDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQES 103

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           |::::      :::|:| :|  :: :::|||: :||||||| :: | :|    |  : ::
Db     104 LAQARLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQL 163

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             | |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::||||||
Db     164 ALDPSLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 223


RESULT 4
AAB51661
ID   AAB51661 standard; Protein; 233 AA.
XX
AC   AAB51661;
XX
DT   26-FEB-2001  (first entry)
XX
DE   Human secreted protein sequence encoded by gene 42 SEQ ID NO:101.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; skin aging; food additive; preservative.
XX
OS   Homo sapiens.
XX
PN   WO200061620-A1.
XX
PD   19-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US09069.
XX
PR   09-APR-1999;   99US-0128702.
PR   20-JAN-2000; 2000US-0177049.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-619225/59.
DR   N-PSDB; AAC93405.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 11; Page 486-487; 540pp; English.
XX
CC   The polynucleotide sequences given in AAC93364 to AAC93412 encode the
CC   human secreted proteins given in AAB51620 to AAB51668. AAB51669 to
CC   AAB51722 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC   neuroprotective; antibacterial; virucide; fungicide; and
CC   ophthalmological. The polynucleotides and polypeptides can be used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. AAC93355 to AAC93363 and
CC   AAB51619 represent sequences which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence   233 AA;

  Query Match             27.0%;  Score 271;  DB 21;  Length 233;
  Best Local Similarity   35.0%;  Pred. No. 2.1e-22;
  Matches   63;  Conservative   45;  Mismatches   66;  Indels    6;  Gaps    4;

Qy       6 SLSSKNPASISELLDCGYHPESLLSDFDYWDYV-VPEPNLNEVIFEESTCQNLVKMLENC 64
           || | :  |: :  : |: ||    |  | | | :|:  |     :|  | ||:::|:
Db      48 SLESSDCESL-DSSNSGFGPE---EDTAYLDGVSLPDFELLSDPEDEHLCANLMQLLQES 103

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           |::::      :::|:| :|  :: :::|||: :||||||| :: | :|    |  : ::
Db     104 LAQARLGSRRPARLLMPSQLVSQVGKELLRLAYSEPCGLRGALLDVCVEQGKSCHSVGQL 163

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
             | |:||||:|||| : ::  |   :  | |    |  || ::| ||:|||::||||||
Db     164 ALDPSLVPTFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 223


RESULT 5
AAY03634
ID   AAY03634 standard; Protein; 229 AA.
XX
AC   AAY03634;
XX
DT   04-JUN-1999  (first entry)
XX
```

DE Hypoxia-regulated gene RTP801 product.
XX
KW Hypoxia-regulated gene; therapeutic; diagnostic; hypoxia; ischemia;
KW apoptosis; angiogenesis; tumorigenic cell; trauma; limb reattachment;
KW revascularisation.
XX
OS Rattus sp.
XX
PN WO9909049-A1.
XX
PD 25-FEB-1999.
XX
PF 21-AUG-1998; 98WO-US17296.
XX
PR 21-AUG-1997; 97US-0056453.
XX
PA (KOHN/) KOHN K I.
PA (QUAR-) QUARK BIOTECH INC.
XX
PI Einat P, Skaliter R;
XX
DR WPI; 1999-180965/15.
DR N-PSDB; AAX29137.
XX
PT New isolated hypoxia-related genes - used to develop products for
PT use in therapy and diagnosis in e.g. hypoxia, ischaemia, apoptosis
PT and angiogenesis
XX
PS Claim 14; Page 81-82; 92pp; English.
XX
CC Sequences AAY03632-36 represent products of hypoxia-regulated genes of
CC the invention. The genes and their products can be used therapeutically
CC and diagnostically in hypoxia, ischemia, apoptosis and angiogenesis. The
CC products and methods can be used for e.g. inducing apoptosis in
CC tumorigenic cells or angiogenesis in trauma situations where e.g. a limb
CC must be reattached or in a transplant where revascularisation is needed.
XX
SQ Sequence 229 AA;

Query Match 26.0%; Score 261; DB 20; Length 229;
Best Local Similarity 30.7%; Pred. No. 2.8e-21;
Matches 65; Conservative 45; Mismatches 68; Indels 34; Gaps 5;

```
Qy       3 ATGSLSSKNPA--------------------SISELLDC--------GYHPESLLSDFD 33
           :: | ||: ||                     :  |  ||        |: ||    |
Db      12 SSSSSSSRTPAADRPPRSAWGSAAREEGLDRCASLESSDCESLDSSNSGFGPE---EDSS 68

Qy      34 YWDYV-VPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDV 92
           | | | :|:  |     :|  | ||:::|:  ||:::      :::|:| :|  :: :::
Db      69 YLDGVSLPDFELLSDPEDEHLCANLMQLLQESLSQARLGSRRPARLLMPSQLLSQVGKEL 128

Qy      93 LRLSSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRD 152
           |||: :|||||| :: | :|    |  : ::  | |:||||:|||| : ::  |   :
Db     129 LRLAYSEPCGLRGALLDVCVEQGKSCHSVAQLALDPSLVPTFQLTLVLRLDSRLWPKIQG 188

Qy     153 FFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
               |      |: ::| ||:|||::||||||
Db     189 LLSSANSSLVPGYSQSLTLSTGFRVIKKKLYS 220
```

RESULT 6
ABB63344
ID ABB63344 standard; Protein; 299 AA.
XX
AC ABB63344;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16824.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07447.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 16824; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 299 AA;

Query Match 21.1%; Score 211.5; DB 22; Length 299;
Best Local Similarity 35.9%; Pred. No. 1.7e-15;
Matches 51; Conservative 24; Mismatches 60; Indels 7; Gaps 3;

```
Qy      43 NLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCG 102
           ||::|    |  : | : |:  |  :|:  | |::| :|  ||||||  :::|:|  ||||
Db     162 NLDDV--SASAVRELSQQLQAQLRDAKRRHLACTEVTLPNDLTQRIAAEIIRMSEREPCG 219

Qy     103 LRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFS 161
           | | : :  | | |  |::     |   |  ||| |  :|:   |:|   |      |
Db     220 ERACTLFIEFESEPNKVKRIAYFKVDPDTVSIFELYLTLRQDKSGWSSLVPQFIKNLTRS 279

Qy     162 SGFRRTLILSSGFRLVKKKLYS 183
           :    |: :|  | | ||||||
Db     280 N----TINISPDFTLTKKKLYS 297
```

RESULT 7
ABB63379
ID ABB63379 standard; Protein; 280 AA.
XX
AC ABB63379;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16929.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.

```
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07482.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 16929; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   280 AA;

  Query Match             19.3%;  Score 193;  DB 22;  Length 280;
  Best Local Similarity   29.2%;  Pred. No. 2e-13;
  Matches   49;  Conservative   36;  Mismatches   61;  Indels   22;  Gaps    5;

Qy      30 SDFDYWDYVVPE----------PNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVL 79
           |: :|: |   |           | ::   || :    :::|:  |  :|   |.|::|
Db      99 SNSNYYYYAADEEEGGSADYALSNYDKKAVEELS----LRLLDE-LRAAKSRHLTCTEVS 153

Qy      80 VPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIE-NVCKKLDRIVCDSSVVPTFELTL 138
           :|  ||  :|::::|:|  || |:||| :::  | |    :::  |  ||  | |||: |
Db     154 LPCDLTPSVAREIIRVSEKEPRGIRGCTIYIEFEDEPKNSRRIASIKVDSDTVSTFEVYL 213

Qy     139 VFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIG 186
             :|::  |||         :|     ||: :|  : : | ||||  |
Db     214 TLRQDHRGWTSLLP------QFMKSLARTITISPEYTITKNKLYSADG 255


RESULT 8
AAB45148
ID   AAB45148 standard; Protein; 74 AA.
XX
AC   AAB45148;
XX
DT   12-FEB-2001  (first entry)
XX
DE   Human secreted protein sequence encoded by gene 29 SEQ ID NO:89.
XX
KW   Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide; cancer;
KW   opthalmalogical; autoimmune disease; hyperproliferative disorder;
KW   cardiovascular disorder; cerebrovascular disorder; wound healing;
KW   nervous system disorder; aging; chemotaxis.
XX
OS   Homo sapiens.
XX
PN   WO200058467-A1.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US07505.
XX
PR   26-MAR-1999;   99US-0126502.
PR   17-DEC-1999;   99US-0172410.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-611712/58.
DR   N-PSDB; AAC80559.
XX
PT   Nucleic acid molecules encoding human secreted proteins, used in
PT   preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT   Parkinson's diseases and cancers -
XX
PS   Claim 11; Page 391-392; 440pp; English.
XX
CC   Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human
CC   secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent
CC   alternative polypeptides encoded by the genes, and amino acid sequences
CC   to which they are homologous. The genes and proteins have activities
CC   dependent on the tissues and cells in which they are expressed. Examples
CC   of their activities include immunosuppressive; antiarthritic;
CC   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
CC   fungicide; and opthalmalogical. The secreted proteins, polynucleotides,
CC   antagonists and agonists may be useful in treating, preventing and/or
CC   diagnosing diseases and disorders such as autoimmune diseases
CC   e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms
CC   of the breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC   cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC   system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC   The polypeptides can also be used to aid wound healing and epithelial
CC   cell proliferation, to prevent skin aging due to sunburn, to maintain
CC   organs before transplantation, for supporting cell culture of primary
CC   tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC   also be used as a food additive or preservative to increase or decrease
CC   storage capabilities. AAC80522-C80530 and AAB45119 represent sequences
CC   used in the isolation and characterisation of the genes and proteins of
CC   the invention.
XX
SQ   Sequence   74 AA;

  Query Match             11.2%;  Score 112.5;  DB 21;  Length 74;
  Best Local Similarity   46.2%;  Pred. No. 4.6e-05;
  Matches   24;  Conservative   10;  Mismatches   17;  Indels    1;  Gaps    1;

Qy     133 TFELTLVFKQENCSWTSFRDFFFS-RGRFSSGFRRTLILSSGFRLVKKKLYS 183
           ||:|||| : ::  |   :  | |    |  || ::| ||:|||::||||||
Db      13 TFQLTLVLRLDSRLWPKIQGLFSSANSPFLPGFSQSLTLSTGFRVIKKKLYS 64


RESULT 9
ABB64251
ID   ABB64251 standard; Protein; 1066 AA.
XX
AC   ABB64251;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 19545.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
```

PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08354.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 19545; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1066 AA;

Query Match 8.5%; Score 85; DB 22; Length 1066;
Best Local Similarity 25.6%; Pred. No. 2.6;
Matches 30; Conservative 21; Mismatches 44; Indels 22; Gaps 4;

```
Qy     36 DYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRL 95
          :|:   |||   :      | || :|      :| || |       | : | : :|| :
Db    648 NYLATVPNLQHFV-----VQALVSLL------AKLTKYGWFDSYKEEMVFQNLLEDVKKF 696

Qy     96 SSTEPCGLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSWTSFRD 152
             |:| | |  : ::     | ::||: : |   :: : | :     |||||
Db    697 -------LQGSVEHCTIGVQ----ILSQLVCEMNSVVEMDVQVSFSKMRKIATSFRD 742
```

RESULT 10
AAB60490
ID AAB60490 standard; Protein; 934 AA.
XX
AC AAB60490;
XX
DT 24-APR-2001 (first entry)
XX
DE Human cell cycle and proliferation protein CCYPR-38, SEQ ID NO:38.
XX
KW Cell cycle and proliferation protein; CCYPR; human; agonist;
KW antagonist; gene therapy; detection; gene therapy;
KW transgenic animal disease model; immune disorder;
KW developmental disorder; cell signalling disorder;
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW arteriosclerosis; asthma; allergy; diabetes mellitus;
KW menstrual cycle disorder; bacterial infection.
XX
OS Homo sapiens.
XX
PN WO200107471-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-US19948.
XX
PR 21-JUL-1999; 99US-0145075.
PR 08-SEP-1999; 99US-0153129.
PR 10-NOV-1999; 99US-0164647.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX
DR WPI; 2001-112727/12.
DR N-PSDB; AAF59627.
XX
PT Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signaling disorders and cell proliferative disorders including cancer -
XX
PS Claim 1; Page 146-148; 205pp; English.
XX
CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoasays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
SQ Sequence 934 AA;

Query Match 8.2%; Score 82; DB 22; Length 934;
Best Local Similarity 25.4%; Pred. No. 4.7;
Matches 53; Conservative 29; Mismatches 75; Indels 52; Gaps 10;

```
Qy     10 KNPASISELLDCGY--------HPESLLSDFDYWDYVVPEPNLNEVI--FEESTCQNLV- 58
          | | :  ||| |            |||||    :|::  || || ::  |   |  ||:
Db     76 KYPNTACELLTCDVPQISDRLGGDESLLSLL--YDFLDHEPPLNPLLASFFSKTIGNLIA 133

Qy     59 KMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDVLRL-SSTEPCGLRGCVMHVNLEIENV 117
          :  |  ::  |:     | ||     |  :   :||| |  || |||  |:|  |  | |
Db    134 RKTEQVITFLKKKDKFISLVL-KHIGTSALMDLLLRLVSCVEPAGLRQDVLHW-LNEEKV 191

Qy    118 CKKLDRIV----------------CD------------SSVVPTFELTLVFKQENCSWT 148
           ::|  ::                 ||             :: |   || :  :::|
Db    192 IQRLVELIHPSQDEDRQSNASQTLCDIVRLGRDQGSQLQEALEPDPLLTALESRQDCVEQ 251

Qy    149 SFRDFFFSRGRFSSGFRRTLILSSGFRLV 177
           :: |        | |    | || :::
Db    252 LLKNMF-------DGDRTESCLVSGTQVL 273
```

RESULT 11
AAB92900
ID AAB92900 standard; Protein; 498 AA.
XX
AC AAB92900;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11520.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

```
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11520; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   498 AA;

  Query Match             8.0%;  Score 80.5;  DB 22;  Length 498;
  Best Local Similarity   23.1%;  Pred. No. 2.9;
  Matches   39;  Conservative   33;  Mismatches   58;  Indels   39;  Gaps    7;

Qy      27 SLLSDFDYWDYVVPE--PNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKL 84
           :|: :   |:  |||    :||:         ||::: |  |: :| :  | :::  ||
Db     292 NLMKNIRDWNTKVPELCLGINELSSHPHNLLWLVQLVPNWTSRGRQLR-QCLSLVIISKL 350

Qy      85 TQRIAQDVLRLSSTEPCGLRGCVMH---VNLEIENVCKKL---------DRIVCDS---- 128
             :||   |:     |:  |:|   | ::  :: ||:          | |: ||
Db     351 LDEKHEDVPNASN-----LQVSVLHRYLVQMKPSDLLKKMVLKKKAEQPDGIIDDSLHLE 405

Qy     129 ----SVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSG 173
               :   |: |  :  : :|| :           |||| |:  :|  |
Db     406 LEKQAYYLTYILLHLVGEVSCSHS----------FSSGQRKHFVLLCG 443


RESULT 12
ABB59933
ID   ABB59933 standard; Protein; 606 AA.
XX
AC   ABB59933;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 6591.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04036.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 6591; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   606 AA;

  Query Match             7.9%;  Score 79.5;  DB 22;  Length 606;
  Best Local Similarity   19.3%;  Pred. No. 4.9;
  Matches   35;  Conservative   32;  Mismatches   47;  Indels   67;  Gaps    9;

Qy      37 YVVPEPNLNEVI------------------FEESTCQNLVKMLENCLSKSKQTKLGCSK 77
           |||   ||:|:                    : :| ||        ||:    :| ||:|
Db     382 YVVSTDNLHEITPEKRQCLFDDERSLRFFRSYSQSNCQT------ECLANYTVSKCGCAK 435

Qy      78 VLVPEKLTQRIAQDVLRLSSTEPCGLR--GCV------MHVNLEIENVCKKLDR---IVC 126
             :|: |             |  |||:   |       ::  :: : : | :|    | |
Db     436 FWMPKPL------------GTPVCGLKDINCYTSAQDELYTLMQNQTMAKSIDESVDITC 483

Qy     127 D-----SSVVPTFELTLVFKQENCSWTSFRD------------FFFSRGRFSSGFRRTL 168
           :     :|:   ||::        :  :||:             :|   :|::  :||:
Db     484 NCMPACTSLEYNFEISRAKYDVAKTIRAFREEYEHTDAIGSRLSVYFKEHQFTA-IKRTI 542

Qy     169 I 169
           :
Db     543 L 543


RESULT 13
AAM79768
ID   AAM79768 standard; Protein; 860 AA.
```

```
XX
AC   AAM79768;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3414.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52901.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 327-328; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   860 AA;

  Query Match             7.9%;  Score 79.5;  DB 22;  Length 860;
  Best Local Similarity   20.4%;  Pred. No. 8;
  Matches   47;  Conservative   41;  Mismatches   79;  Indels   63;  Gaps   10;

Qy       2 VATGSL------SSKNPASISELLDCGYHPESLLSDF--------DYWDYVVPEPNLNEV 47
           : || |       :|:   :|    :  : || ::  :          | | :|   :| :
Db     646 IKTGELLGFPVYASRKSIAIRNXNNDKWPPERIIQYYGPATXAQDGSWGYRIPIYMINRI 705

Qy      48 IFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQ--------RIAQDVLRLSSTE 99
           |      | ::|::         |  | :  :: :: ||        |:| | |  :  |
Db     706 I----RLQAVLKII---------TATGRALTILAQQETQMRNAIYQNRLALDYLLAAEGE 752

Qy     100 PC---GLRGCVMHVNLEIENVCKKLDRIVCDSSVVPTFELTLVFKQENCSW------TSF 150
           |    |  | :|    |:|  : :: || | : |    : :        |        |
Db     753 VCRKFNLTNCCLH----IDNQGQVVEDIVRDMTKVAHVPVQV--------WHGFDPGAMF 800

Qy     151 RDFFFSRGRFSSGFRRTLILSSGFRLVKKKL-------YSLIGTTVIEGS 193
           | :| : | | :   | :|:   : |: : |        | | | | : :
Db     801 RKWFPALGGFKTLIIRVIIVIGTYLLLPRLLPVLLQMIKSFIATLVYQNA 850


RESULT 14
AAU20460
ID   AAU20460 standard; Protein; 169 AA.
XX
AC   AAU20460;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human secreted protein, Seq ID No 452.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing.
XX
OS   Homo sapiens.
XX
PN   WO200155326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451931/48.
DR   N-PSDB; AAS33169.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Claim 11; SEQ ID No 452; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
CC   angina and thrombosis), infections caused by bacteria, viruses and
CC   fungi and ocular disorders (e.g. corneal infections). (I) and (II),
```

```
CC   agonists, antagonists and antibodies can also be used to promote wound
CC   healing, maintain organs before transplantation, and support cell culture
CC   of primary tissues. AAU20342-AAU20666 represent human secreted protein
CC   amino acid sequences, and related sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO
CC   at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   169 AA;

  Query Match             7.9%;  Score 79;  DB 22;  Length 169;
  Best Local Similarity   23.4%;  Pred. No. 0.95;
  Matches   41;  Conservative   16;  Mismatches   50;  Indels   68;  Gaps    8;

Qy      33 DYWDYVVPEPNLNEVIFEESTCQNLVKMLENCLSKSKQTKLGCSKVLVPEKLTQRIAQDV 92
           |||            ::    |            ::| |          :  | |  ::|
Db       7 DYW-----------LYSRGVC-----------RTKSTS---------DNETSRKNEEV 33

Qy      93 LRLSST-EPCGLRGCVMHVNLEIENVCKKLD----------------RIVCDSSVVPTF 134
           :  |     | | |           ||||:|                | |  ||| |
Db      34 MTHSGLWRTCCLEGA-------FRGVCKKIDHFPEDADYEQDTAEYLLRAVRASSVFPIL 86

Qy     135 ELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSLIGTTV 189
            :||:|    |   |  :|  |        |  :|||:|   |   | ::||  |
Db      87 SVTLLFFGGLCVAAS--EFHRS--------RHNVILSAGIFFVSAGLSNIIGIIV 131


RESULT 15
AAR53277
ID   AAR53277 standard; Protein; 261 AA.
XX
AC   AAR53277;
XX
DT   23-JUN-1994  (first entry)
XX
DE   Ced-4 (Q262X).
XX
KW   C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW   embryogenesis; cell death; hydrophilic; transmembrane; region;
KW   hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW   protein synthesis; null phenotype; calcium-binding domain.
XX
OS   Caenorhabditis elegans.
XX
PN   WO9325685-A.
XX
PD   23-DEC-1993.
XX
PF   14-JUN-1993;   93WO-US05701.
XX
PR   12-JUN-1992;   92US-0897788.
PR   20-NOV-1992;   92US-0979638.
XX
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Horvitz HR,  Shaham S,  Yuan J;
XX
DR   WPI; 1994-007542/01.
DR   N-PSDB; AAQ64641.
XX
PT   Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT   develop agents to increase or prevent cell death in organisms
XX
PS   Claim 10; Fig 1; 127pp; English.
XX
CC   The sequences given in AAR53274-78 are encoded by mutant versions of the
CC   C. elegans ced-4 gene.  A 2.2 kb mRNA was identified as the wild type
CC   ced-4 transcript.  This transcript was present at normal levels in a
CC   ced-3 mutant, suggesting that ced-3 is not a transcriptional regulator
CC   of ced-4.  The 2.2 kb transcript is primarily expressed during embryo-
CC   genesis, which is consistent with observed cell deaths in C. elegans.
CC   The wild type Ced-4 protein is 549 amino acids in length and has a
CC   relative molecular mass of 62,877.  Ced-4 is highly hydrophilic with
CC   a predicted pI of 5.12.  There are no obvious transmembrane regions.
CC   The longest hydrophobic region is a segment of 12 amino acids from
CC   residue 382-393.  Of the eight mutations which occur within the ced-4
CC   gene one results in a single amino acid substitution and the other
CC   seven appear to prevent either ced-4 RNA splicing or completion of
CC   Ced-4 protein synthesis.  These seven mutations establish the null
CC   phenotype of the ced-4 gene, confirming that ced-4 function is not
CC   essential for viabillity.  Ced-4 contains two regions which have
CC   similarity with known calcium-binding domains, suggesting that Ced-4
CC   activity may be modulated by calcium.
XX
SQ   Sequence   261 AA;

  Query Match             7.9%;  Score 79;  DB 15;  Length 261;
  Best Local Similarity   24.7%;  Pred. No. 1.7;
  Matches   45;  Conservative   23;  Mismatches   46;  Indels   68;  Gaps   11;

Qy      10 KNPASISELLD-CGYHPESLLSDF--DYWDYVVPEPNLNE--VIFEESTCQNLVKMLENC 64
           :  : :  |:|   |: :| |:||  || |: : ||:|    ||  : : | |
Db      63 RQASELGPLIDFFNYNNQSHLADFLEDYIDFAINEPDLLRPVVIAPQFSRQML------- 115

Qy      65 LSKSKQTKLGCSKVLVPEKLTQRIAQDVLRLSSTEPCGLRGCVMHVNLEIENVCKKLDRI 124
           ::  ||     ||:::|                | :|        :: | ||||
Db     116 ---DRKLLLG----NVPKQMT--------------CYIR------EYHVDRVIKKLD-- 145

Qy     125 VCDSSVVPTFELTLVFKQENCSWTSFRDFFFSRGRFSSGFRRTLILSSGFRLVKKKLYSL 184
                       | |   ||  | |  ||  ||  :::| |        |   |
Db     146 -----------------EMCDLDSF--FLFLHGRAGSG--KSVIASQAL----SKSDQL 179

Qy     185 IG 186
           ||
Db     180 IG 181


Search completed: October 11, 2002, 14:11:14
Job time : 64 secs
```